
Bachelorarbeit

Herr
Eric Zuchantke

Entfaltung von Membranproteinen auf der Basis von Monte-Carlo- Simulationen

Mittweida, 2012

Bachelorarbeit

Entfaltung von Membranproteinen auf der Basis von Monte-Carlo- Simulationen

Autor:
Herr

Eric Zuchantke

Studiengang:
Biotechnologie/Bioinformatik

Seminargruppe:
BI09w3-B

Erstprüfer:
Prof. Dr. rer. nat. Dirk Labudde

Zweitprüfer:
Dipl.-Inf. (FH) Daniel Stockmann

Einreichung:
Mittweida, 13.08.2012

Verteidigung/Bewertung:
Mittweida, 2012

Bibliographische Beschreibung:

Zuchantke, Entfaltung von Membranproteinen auf der Basis von Monte-Carlo-Simulationen - 2012. – 11, 70, 1 S. Mittweida, Hochschule Mittweida, Fakultät Mathematik/Naturwissenschaften/Informatik, Bachelorarbeit, 2012

Englischer Titel

Unfolding of membrane proteins on the basis of Monte-Carlo simulations

Kurzbeschreibung

In dieser Arbeit sollte die Entfaltung von Membranproteinen bzw. Abschnitten dieser simuliert werden. Dazu wurde ein Programmpaket geschaffen, mit dem sich Proteine mittels des Branch-and-Bound Algorithmus modellieren lassen und über Monte-Carlo Verfahren entfalten lassen. Dabei ist im Vergleich zu anderen Simulation bzw. physischen Experimenten, die Konformation der betrachteten Struktur immer bekannt und kann einem Kraft-Wert sowie einem End-End Abstand eindeutig zugeordnet werden. Als Test für die Simulation wurde die Helix A, B und C des Bacteriorhodopsin entfaltet.

Danksagung

Ich möchte mich bei meinen Betreuern Prof. Dirk Labudde und Dipl.-Inf. Daniel Stockmann für ihre Unterstützung und Ratschläge während der Arbeit und auch über das gesamte Studium hinaus bedanken.

Desweiteren möchte ich mich bei B.Sc. Florian Heinke bedanken, welcher mit seinen Ratschlägen oftmals kleinere und größere Probleme beseitigen konnte und viele Anregungen gab.

Natürlich möchte ich auch Susanne Bause, Christin Lembke und Julia Mitev danken, welche mir mit Rat und Tat zur Seite standen, was die Formulierung und das Aussehen der Arbeit betrifft.

Zum Schluss gilt mein Dank noch meiner Familie welche mich über das ganze Studium unterstützt haben und dieses mir mit ihrer Hilfe erst ermöglichten.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	III
Tabellenverzeichnis	VI
Abkürzungsverzeichnis	VII
1 Einleitung.....	1
2. Proteine.....	3
2.1 Membranproteine	3
2.2 Aufbau und Struktur.....	6
2.2.1 Primärstruktur	9
2.2.2 Sekundärstruktur.....	9
2.2.3 Tertiär- und Quartärstruktur	11
2.3 Faltung und Entfaltung von Proteinen	12
3 Energieprofile und Energielandschaften	16
3.1 Aussagen von Energieprofilen und Energielandschaften	16
3.2 Energiefunktion.....	18
4 Einzelmolekülkraftspektroskopie	22
5 Simulation.....	27
5.1 Fragment-Bibliothek	27
5.2 Modellierung mittels Branch and Bound	31
5.3 Bestimmung einzelner Abschnitte	36
5.4 Entfaltungssimulation.....	38
6. Entfaltung von Bacteriorhodopsin Helix A,B und C	44
6.1 Ergebnisse Helix A	49

6.2 Ergebnisse Helix B	52
6.2.1 Entfaltung mit statischem Teil	52
6.2.2 Entfaltung Helix B ohne statischen Teil.....	55
6.3 Ergebnisse Helix C	60
6.2.1 Entfaltung mit statischem Teil	60
6.2.2 Entfaltung von Helix C ohne statischen Teil.....	64
7 Auswertung	68
8 Zusammenfassung und Ausblick.....	70
Literaturverzeichnis	VIII
Anhang.....	XII
Selbstständigkeitserklärung	XIII

Abbildungsverzeichnis

Abbildung 1: Darstellung grundlegender Strukturbetrachtung für Simulation	2
Abbildung 2: Darstellung verschiedener integraler Membranproteine	3
Abbildung 3: Darstellung verschiedener peripherer Membranproteine	4
Abbildung 4: allgemeine Form einer Aminosäure	7
Abbildung 5: Venn Diagramm der 20 Aminosäuren	7
Abbildung 6: Strukturschichten eines Protein im Überblick	8
Abbildung 7: Ramachandran-Plot	10
Abbildung 8: Aufbau Sekundärstrukturen	11
Abbildung 9: Two-State-Model	14
Abbildung 10: Darstellung einer Energielandschaft	17
Abbildung 11: Übersicht über Single Molecule-Methoden	22
Abbildung 12: Darstellung SMFS mittels AFM	23
Abbildung 13: Kraft-Abstandskurve mittels Force-Clamp Methode	24
Abbildung 14: Kraft-Abstandskurve mittels Velocity-Clamp Methode	25
Abbildung 15: schematische Darstellung WLC	26
Abbildung 16: Übersicht über erstellte Klassen	27
Abbildung 17: Aufbau eines Fragmentes	28
Abbildung 18: schematischer Ablauf Clustering	29
Abbildung 19: Häufigkeitsverteilung der Anzahl an Fragment-Schwerpunkten	30
Abbildung 20: Klasse Protein_Fragment	30
Abbildung 21: Klassendiagramm für das Verbinden von Fragmenten	33
Abbildung 22: Klassendiagramm für die Klasse Entfaltung	41
Abbildung 23: <i>Halobacterium salinarum</i>	44

Abbildung 24: Darstellung 1BRR.....	44
Abbildung 25: Helix A-B-C im nativen Zustand	45
Abbildung 26: schematischer Aufbau der Helices mittels Fragment-Namen....	46
Abbildung 27: Kraft-Abstandskurve der Helix A	49
Abbildung 28: Energetischer Entwicklung der Entfaltung von Helix A.....	50
Abbildung 29: Kraft über Entfaltungsdauer der Helix A mit jeweiligen Strukturen	50
Abbildung 30: Kraft-Abstandskurven für Helix B mit statischen Teil.....	52
Abbildung 31: Energetische Entwicklung der Entfaltung von Helix B mit statischem Teil.....	53
Abbildung 32: Kraftverlauf über die Entfaltungsdauer von Helix B mit statischen Teil.....	54
Abbildung 33: Kraft-Abstandskurve für Helix B ohne statischen Teil.....	56
Abbildung 34: Energetische Entwicklung der Entfaltung von Helix B ohne statischen Teil.....	56
Abbildung 35: Energieprofile der intermediären Form der Helix B ohne statischen Teil.....	57
Abbildung 36: Kraftverlauf über die Entfaltungsdauer von Helix B ohne statischen Teil.....	58
Abbildung 37: Kraft-Abstandskurve Helix C mit statischen Teil.....	60
Abbildung 38: Energetischer Entwicklung der Entfaltung von Helix C mit statischem Teil.....	61
Abbildung 39: Energieprofil der intermediären Helix C Struktur aller 3 Verläufe	61
Abbildung 40: Kraftverlauf über die Entfaltungsdauer der Helix C mit statischem Teil.....	62
Abbildung 41: Darstellung unmöglicher Entfaltung der Helix C	63

Abbildung 42: Kraft-Abstandskurve der Helix C ohne statischen Teil	64
Abbildung 43: Energetischer Entwicklung der Entfaltung von Helix C ohne statischen Teil.....	65
Abbildung 44: Kraftverlauf über die Entfaltungsdauer mit Strukturen der Helix C ohne statischen Teil.....	66
Abbildung 45: Kraft-Abstandskurven der Helix A, B und C	68

Tabellenverzeichnis

Tabelle 1: 20 proteinogene Aminosäuren mit Namen	6
Tabelle 2: Häufigkeitstabelle mit Innen/Außen-Zuständen	19
Tabelle 3: Wahrscheinlichkeit p_i	20
Tabelle 4: Ergebnis der TMDet Abfrage zu 1BRR	36
Tabelle 5: Übersicht Anzahl gebildeter Strukturen mittels Branch and Bound .	46
Tabelle 6: Vergleich Anzahl Strukturen in Abhängigkeit vom statischen Teil ...	47
Tabelle 7: Vergleich der Anzahl an Konformation in Abhängigkeit der Methode <i>verringierung_anzahl()</i>	48
Tabelle 8: Überblick über verwendete Parameter für Entfaltungssimulation	48
Tabelle 9: Vergleich RMSD-Werte für die Konformationen in Helix A	51
Tabelle 10: Häufigkeit der aufgetretenen Konformationen pro Verlauf in Helix A	51
Tabelle 11: RMSD-Werte der intermediären Strukturen der Helix B	54
Tabelle 12: Häufigkeit der Konformationen pro Verlauf für Helix B mit statischen Teil	55
Tabelle 13: RMSD-Wert der auftretenden Strukturen bei Helix B ohne statischen Teil	59
Tabelle 14: Häufigkeit der Konformationen in allen Verläufen von Helix B ohne statischen Teil	59
Tabelle 15: Häufigkeiten der Konformation pro Verlauf für Helix C mit statischem Teil	63
Tabelle 16: RMSD-Werte der Strukturen für Helix C ohne statischen Teil	67
Tabelle 17: Häufigkeit der Konformationen pro Zeitschritt in allen drei Verläufen	67

Abkürzungsverzeichnis

AFM	Atom Force Microscope (Atomkraftmikroskop)
C-Terminus	Carboxylgruppe
N-Terminus	Aminogruppe
f-d-Kurve	Kraft-Abstandskurve
MC	Monte-Carlo
MD	molekular dynamisch
nTM	Nichtransmembranerbereich
RMSD	Root-Mean-Square Deviation
SM	Single Molecule
SMFS	Single Molecule Force Spectroscopy
TM	Transmembranerbereich
WLC	Wormlike Chain

1 Einleitung

Proteine gewinnen ihre Funktion durch ihre Struktur bzw. begrenzt die Struktur die Funktion eines Proteins. Das bedeutet, ein Protein welches nicht in seiner korrekten Form vorliegt, wird nicht die Aufgabe erfüllen, für welche das Protein vorgesehen ist oder die Funktionalität ist stark durch die falsche Struktur beeinträchtigt.

Doch wie kommt es dazu, dass ein Protein nicht seine korrekte Form erhält?

Um diese Frage zu klären und den Grund für die Fehlfaltung von Proteinen zu klären führt man seit nun ca. 20 Jahren sogenannte Einzelmolekülkraftspektroskopien durch. Mit Hilfe dieser können die strukturellen Informationen, während des Prozesses der Entstehung, der Faltung, bzw. der Inaktivierung, der Entfaltung, gewonnen werden. Im Laufe der Zeit gab es immer mehr Erkenntnisse durch diese Experimente, was die mechanische Entfaltung bzw. den Faltungsweg von einzelnen Proteinen angeht. Neben diesen Erkenntnissen konnte der experimentelle Aufbau und die Durchführung verbessert werden, um somit mehr und genauere Informationen zu erhalten [Schlierf & Rief, 2009].

Neben diesen physischen Experimenten werden auch Simulationen durchgeführt. Der Grund für solche Simulationen liegt darin, zu versuchen die vorangestellten Experimente nachzuvollziehen bzw. schon vor einem solchen Entfaltungsexperiment den Ausgang dessen vorherzusagen. Durch diese Vorhersage kann später auf die, im Vergleich zur Simulation teuren und aufwendigen Experimente verzichtet oder diese auch auf Grundlage der Simulationen optimiert werden. Weiterer Vorteil einer solchen Simulation im Vergleich zu den physischen Experimenten ist die hohe Reproduzierbarkeit.

Die Art der Simulationen für Entfaltungssimulation, beschränkt sich dabei auf 2 Formen. Den molekular dynamischen (MD) und den Monte Carlo (MC) Simulationen. Die Auswahl, ob nun MD oder MC Simulationen hängt dabei von der Genauigkeit ab, mit welcher das Molekül betrachtet wird.

Die Genauigkeit, mit welcher ein Molekül in Simulationen betrachtet wird, lässt sich dabei grob in drei Betrachtungsweisen einteilen: Lattice-Model, Off-Lattice-Model und das All-Atom-Model. Abbildung 1 zeigt die Betrachtungsweise der Modelle. Dabei steigt mit der Genauigkeit gleichzeitig der Rechenaufwand für die Simulation [Zhang et al, 2009].

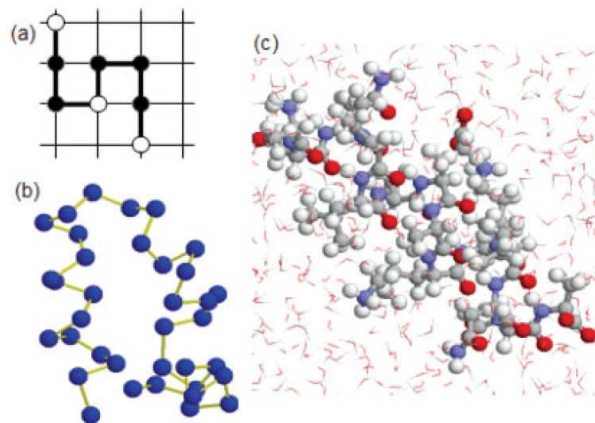


Abbildung 1: Darstellung grundlegender Strukturbetrachtung für Simulation, dabei zeigt (a) Lattice-Model, (b) Off-Lattice-Model und (c) das All-Atom-Model [Zhang et al, 2009].

In dieser Arbeit soll eine Entfaltungssimulation auf Grundlage von MC-Simulationen erstellt werden. Dabei werden die zu entfaltenden Proteine mit dem Off-Lattice-Model betrachtet. Neben der Erstellung von Kraft-Abstandskurven, wie in anderen Simulationen mittels MC-Verfahren [King et al, 2010, Makarov et al, 2001], soll diese Simulation auch die einzelnen intermediären Formen während der Entfaltung darstellen.

2. Proteine

Proteine sind lineare Makromoleküle, welche diverse Stoffwechselreaktionen realisieren und sie somit unabhkmmlich für jeden lebenden Organismus macht. Bei Proteinen wird grundsätzlich zwischen globulären Proteinen und Membranproteinen unterschieden. Unter einem globulären Protein versteht man ein Protein welches, frei im Cytoplasma oder anderen Zellorganellen vorhanden ist [Heinke, 2010]. Diese Proteine besitzen eine kugelähnliche Form und eine ihrer prägendsten Eigenschaften ist ihre gute Wasserlöslichkeit. Das Achsenverhältnis eines solchen Proteins ist kleiner als 10:1 und durchschnittlich kleiner als 4:1 [Löffler et al, 2007].

2.1 Membranproteine

Diese Gruppe von Proteinen kann man in 2 strukturelle Gruppen einteilen, die integralen und peripheren Membranproteine. Integrale Proteine sind in die Membran eingebettet und können dabei auch vom intrazellulären Raum durch die Membran, bis zum extrazellulären Raum verlaufen, siehe Abbildung 2.

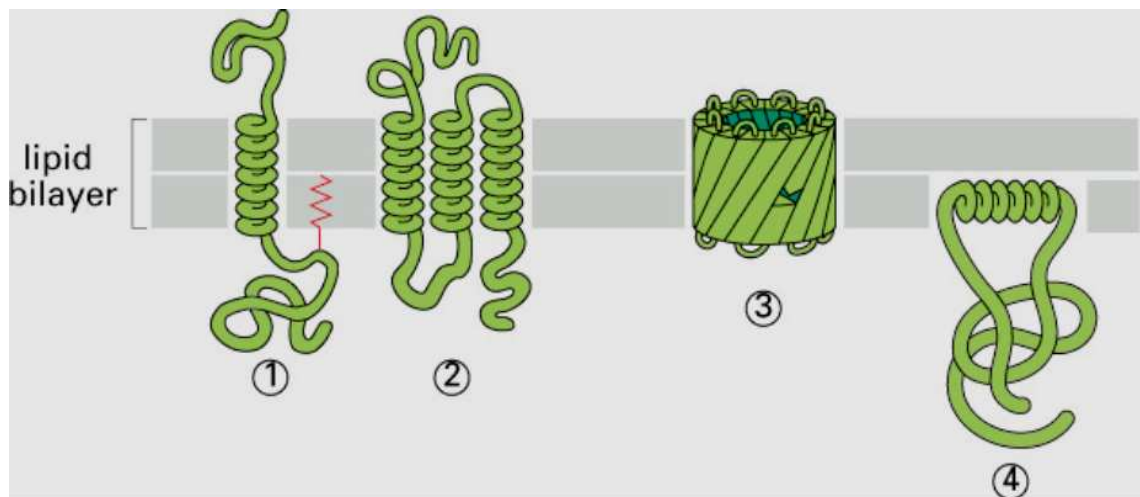


Abbildung 2: Darstellung verschiedener integraler Membranproteine. 1 Einpfad-Transmembranprotein mit zusätzlicher Bindung durch Fettsäurekette, 2 Mehrpfad-Transmembranprotein, 3 Transmembranprotein, 4 Verankerung über Helix auf einer Membranseite [URL 1]

Periphere Proteine hingegen befinden sich nur außerhalb der Membran und treten nur in hydrophobe Wechselwirkung mit anderen Membranproteinen. Sie selbst besitzen keinerlei Kontakt zur hydrophoben Schicht der Membran, dargestellt in Abbildung 3 [URL 1].

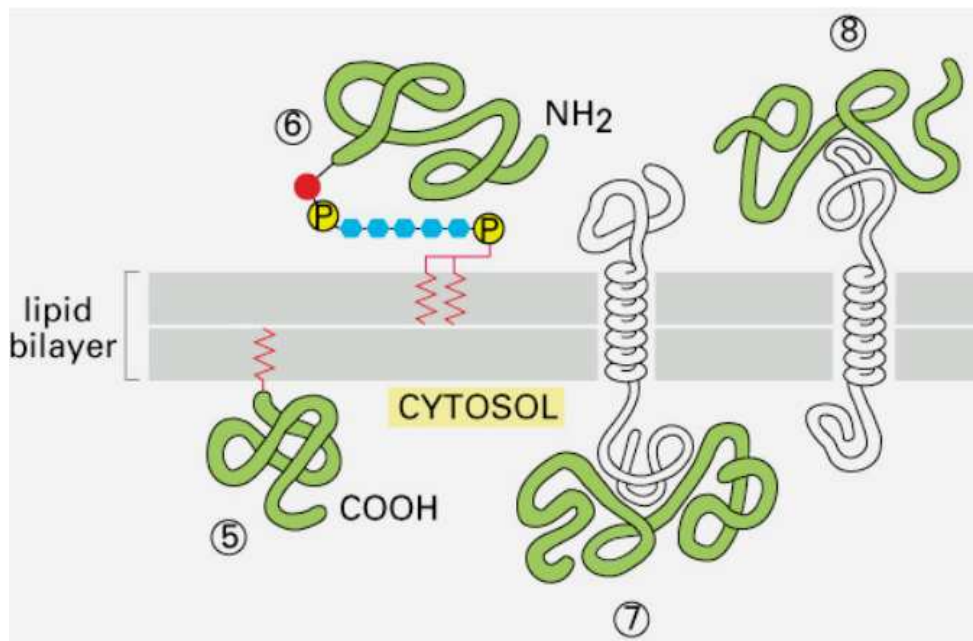


Abbildung 3: Darstellung verschiedener peripherer Membranproteine. 5 verankert über kovalent gebundene Lipidgruppe, 6 verknüpft über Oligosaccharidketten, 7 und 8 Assoziation mit anderen Membranproteinen durch nicht kovalente Wechselwirkungen [URL 1]

Während globuläre Proteine hydrophil sind, sind die Teile des Membranprotein, welche innerhalb der Membran verankert sind hydrophob. Mittels des Hydrophobizitäts-Index, welcher aus der Analyse der Aminosäuresequenz erstellt wird, kann man vorhersagen, welche Teile des Proteins innerhalb der Membran und außerhalb der Membran liegen.

Ein Membranprotein kann hierbei die Funktion eines Rezeptors, Ionenkanals oder Transporters annehmen. Das Membranprotein ist für die Kommunikation, sowie den Stoffwchelaustausch zwischen Intra- und Extrazellulärenraum zuständig. Eine weitere Funktion ist die Zelladhäsion, wobei das Protein sich mittels seiner extrazellulären Domäne mit anderen Zellen oder Strukturproteinen einer extrazellulären Matrix verbinden kann [URL 1].

Diese Eigenschaften machen die Membranproteine interessant für die Pharmazie und Virologie. Das Problem hierbei ist, dass nur wenige Strukturen von Membranproteinen mittels NMR oder Kristallröntgenspektroskopie ermittelt werden konnten, da für analytische Methoden das Protein aus der Membran gelöst werden müsste. Durch die Lösung aus der Membran würde sich aber auch seine Struktur durch die geänderten Umgebungsbedingungen schlagartig ändern [Heinke, 2010].

2.2 Aufbau und Struktur

Ein Protein besteht aus einer Verkettung der 20 proteinogenen Aminosäuren. Diese sind in Tabelle 1 aufgelistet.

Tabelle 1: 20 proteinogene Aminosäuren mit Namen, 1 Letter-Code und 3-Letter-Code

Name	3-Letter-Code	1-Letter-Code	Name	3-Letter-Code	1-Letter-Code
Alanin	Ala	A	Leucin	Leu	L
Arginin	Arg	R	Lysin	Lys	K
Asparagin	Asn	N	Methionin	Met	M
Asparaginsäure	Asp	D	Phenylalanin	Phe	F
Cystein	Cys	C	Prolin	Pro	P
Glutamin	Gln	Q	Serin	Ser	S
Glutaminsäure	Glu	E	Threonin	Thr	T
Glycin	Gly	G	Tryptophan	Trp	W
Histidin	His	H	Tyrosin	Tyr	Y
Isoleucin	Ile	I	Valin	Val	V

Bei diesen 20 Aminosäuren handelt es sich um α -Aminosäuren, deren allgemeiner Aufbau aus einer Carboxylgruppe (C-Terminus) und einer primären Aminogruppe (N-Terminus). Diese beiden funktionellen Gruppen sind Substituenten an dem sogenannten C_{α} -Atom. Neben diesen beiden Gruppen befinden sich an dem α -Kohlenstoff noch ein Wasserstoff-Atom, sowie eine spezifische Seitenkette, welche sich von Aminosäure zu Aminosäure unterscheidet. Die einzige Ausnahme bezüglich des Aufbaus bildet hierbei Prolin. Der allgemeine Aufbau einer Aminosäure ist in Abbildung 4 dargestellt.

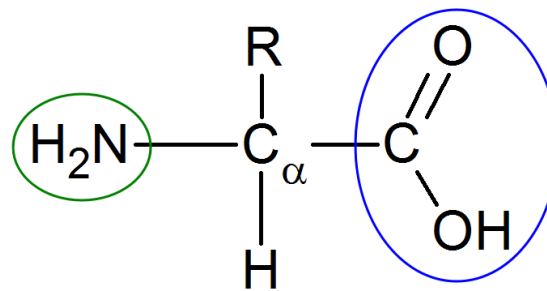


Abbildung 4: allgemeine Form einer Aminosäure, grün umrahmt der N-Terminus und blau umrahmt der C-Terminus

Durch die spezifischen Seitenketten besitzen die Aminosäuren unterschiedliche physiko-chemische Eigenschaften, wie in Abbildung 5 dargestellt.

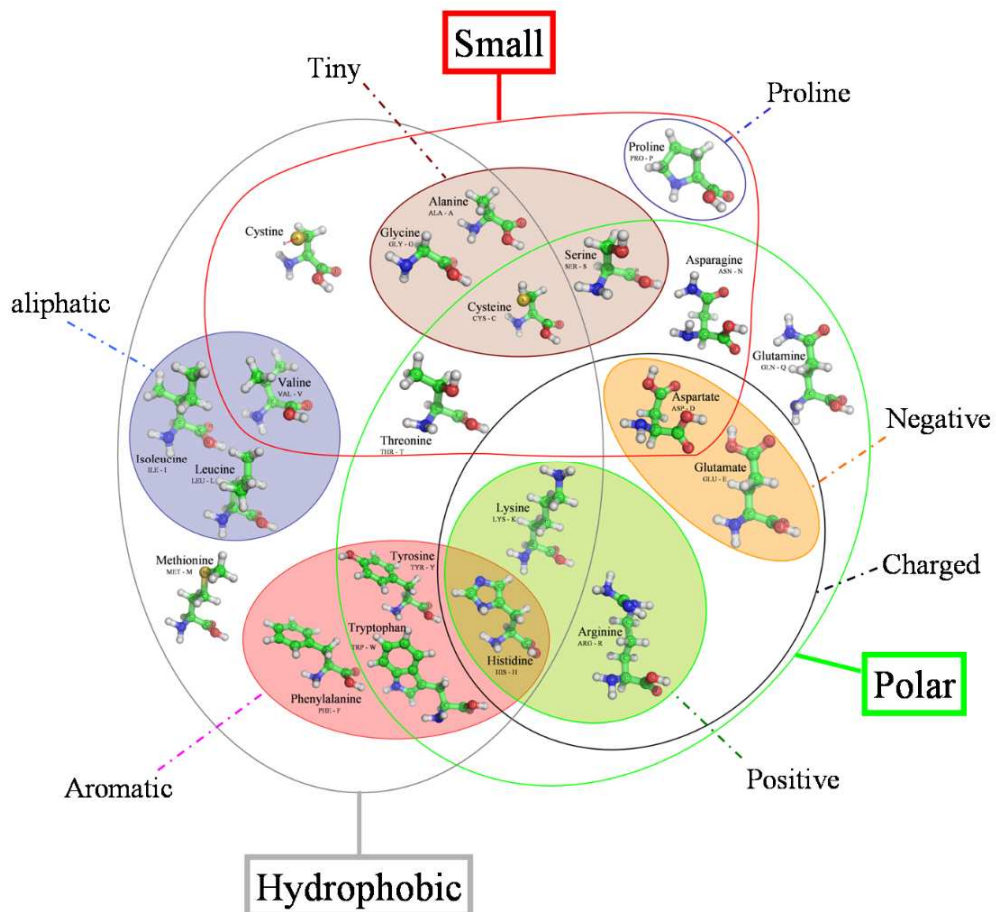


Abbildung 5: Venn Diagramm der 20 Aminosäuren. Dargestellt sind Struktur der Aminosäuren sowie deren Zuteilung zu den Eigenschaften [URL 2]

Die Aminosäuren sind in einer Polypeptidkette durch die sogenannte Peptidbindung miteinander verbunden. Diese Peptidbindung bildet sich unter einer Kondensationsreaktion aus, bei welcher sich der C-Terminus der ersten Aminosäure mit dem N-Terminus der zweiten Aminosäuren verbindet und hierbei Wasser abgespalten wird [Voet et al, 2002].

Durch den spezifischen Aufbau einer Aminosäure und der Verknüpfung mehrerer Proteine miteinander, ergeben sich durch die Lage der einzelnen Atome im Raum und den Bindungswinkeln zwischen den Aminosäuren unterschiedliche Strukturen. Diese räumlichen Strukturen verleihen den Proteinen ihre spezifische Funktion.

Zur Untersuchung von Proteinen kann man unterschiedliche strukturelle Ebenen betrachten. Bei diesen Ebenen handelt es sich um die Primär-, Sekundär-, Tertiär- und Quartärstruktur, dargestellt in Abbildung 6. Dabei steigt, beginnend bei der Primärstruktur bis zur Quartärstruktur, der Informationsgehalt und die Komplexität der jeweiligen Strukturschicht.

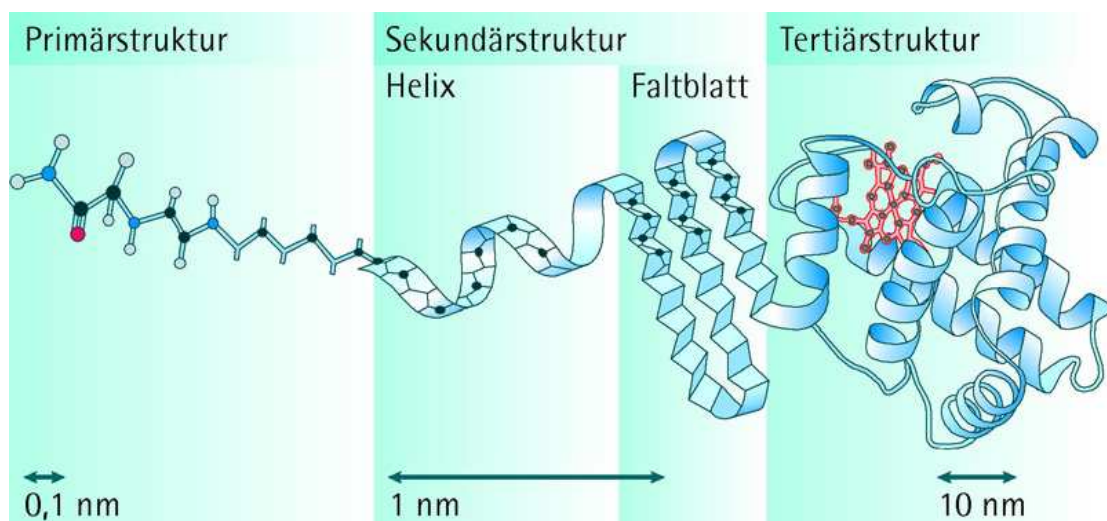


Abbildung 6: Strukturschichten eines Protein im Überblick, v.l.n.r Primärstruktur, Sekundärstruktur mit Helix und Faltblatt, Tertiärstruktur. (Quartärstruktur ist nicht abgebildet) [Schildbach, 2010]

2.2.1 Primärstruktur

Bei der Primärstruktur handelt es sich rein um die Abfolge der Aminosäuren. Es werden keinerlei Informationen zur räumlichen Struktur gegeben. Mittels dieser Betrachtung lassen sich Proteine hinsichtlich ihrer Ähnlichkeit zu anderen Proteinen untersuchen, da verschiedene Aminosäuren ähnliche physikochemische Eigenschaften besitzen (Abbildung 5). Durch den Vergleich der Aminosäuresequenz, können hochkonservierten Bereiche gefunden werden. Unter hochkonservierten Bereichen versteht man Sequenzabschnitte, welche in mehreren Organismen bzw. Sequenz identisch oder ähnlich auftreten. Durch das Vergleichen von Sequenzen, können evolutionäre Entwicklungen bestimmt werden [Voet et al, 2002].

2.2.2 Sekundärstruktur

Bei der Sekundärstruktur werden räumliche Informationen zu den reinen Sequenzinformationen hinzugefügt. Dabei werden einzelne Sequenzbereiche betrachtet und in eine von 3 Sekundärstrukturen kategorisiert, α -Helices, β -Faltblätter und Coil.

Durch die Torsionswinkel, den ϕ - und ψ -Winkel, jeder Aminosäure wird die Konformation der einzelnen Aminosäuren beschrieben. Dabei ist festzustellen, dass die meisten Aminosäuren in der Trans-Konformation vorliegen. Der ϕ -Winkel beschreibt den Winkel zwischen dem C_α -Atom und dem N-Terminus. Mittels des ψ -Winkel wird der Winkel zwischen dem C_α -Atom und dem C-Terminus beschrieben. Durch die spezifischen Seitenketten einer jeden Aminosäure, kann nicht jeder theoretische Winkel praktisch vorkommen. Dadurch ergibt sich, dass nur geringe Bereiche existieren, in welchen die beiden Winkel zulässig sind. Diese Bereiche können mittels eines Ramachandran-Diagrammes beschrieben werden, wie in Abbildung 7 gezeigt. Durch diese Einschränkung ergeben sich für die Sekundärstrukturelemente spezifische Richtlinien, was ihren Aufbau betrifft [Voet et al, 2002].

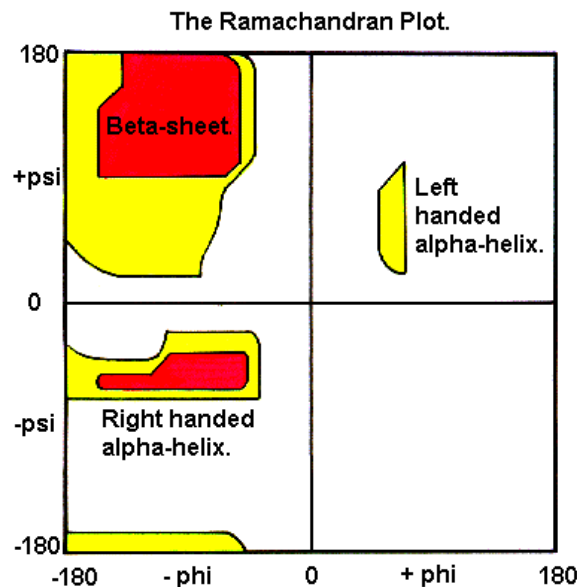


Abbildung 7: Ramachandran-Plot, zeigt zulässige Winkel einer Aminosäure für ein Faltblatt oder Helix [URL 3]

Unter einer α -Helix, dargestellt in Abbildung 8 (1), versteht man eine rechtsgängige Helix, in welcher 3,6 Aminosäuren pro Windung verbaut sind. Im Durchschnitt besteht eine Helix aus 12 Aminosäuren und somit ca. 3 Windungen. Zwischen den einzelnen Windungen bilden sich Wasserstoffbrückenbindungen aus, welches die Stabilität der Helix verstärkt. Die Stabilität einer Helix wird auch durch ihre kompakte Anordnung und den damit verbundenen Van-der-Waals-Kontakten zwischen den Resten innerhalb der Helix verbessert. Der Grund für die Kompaktheit einer Helix besteht darin, dass die einzelnen Seitenkette entgegen der Helixrichtung herausragen [Voet et al, 2002].

Wie auch bei den α -Helices werden die β -Faltblätter, gezeigt in Abbildung 8 (2), durch Wasserstoffbrückenbindungen stabilisiert. Diese Bindungen bilden sich jedoch nicht innerhalb eines Polypeptidstranges, sondern zwischen benachbarten Strängen. Durch die Ausbildung von Wasserstoffbrückenbindungen werden die β -Faltblätter durch die auftretenden Kräfte gestaucht. β -Faltblätter bestehen durchschnittlich aus 6 Polypeptidsträngen. Ein Polypeptidstrang besteht aus durchschnittlich 6 und maximal 15 Aminosäuren. Es gibt 2 Arten von β -Faltblättern, das parallele β -Faltblatt, bei welchem die verbundenen Stränge in dieselbe Richtung verlaufen

und das antiparallele β -Faltblatt, welches in die entgegengesetzte Richtung verläuft [Voet et al, 2002].

Bei der Coil-Struktur handelt es sich um regellose Strukturen. Diese treten zwischen den α -Helices und β -Faltblättern auf. Unter regellos wird hierbei verstanden, dass der Grund für die Bildung dieses Sekundärstrukturelement noch nicht geklärt ist [Voet et al, 2002].

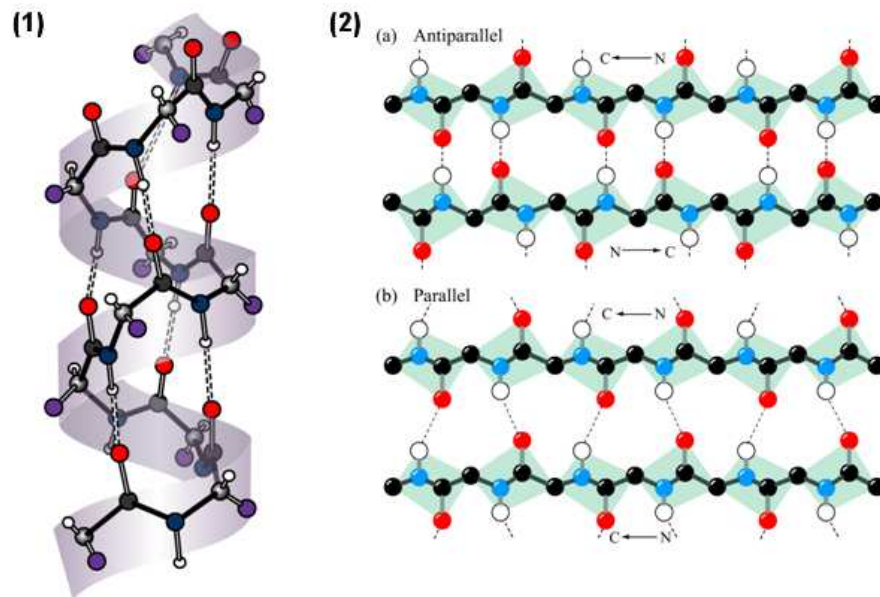


Abbildung 8: Aufbau Sekundärstrukturen, α -Helix (1) und β -Faltblatt (2), (a) zeigt den Aufbau eines antiparallelen und (b) den Aufbau eines parallelen β -Faltblattes [Voet et al, 2002]

2.2.3 Tertiär- und Quartärstruktur

Der Begriff Tertiärstruktur beschreibt die gesamte räumliche Struktur eines Proteins und damit die Anordnung dessen Sekundärstrukturelemente. Dabei sind sämtliche Positionen der einzelnen Atome jeder Aminosäure bekannt [Voet et al, 2002].

Bei Quartärstrukturen handelt es sich um Zusammenschlüsse von mehreren kleinen Proteinen zu einem funktionsfähigen Kollektiv, wobei die einzelnen Proteine als Untereinheiten betrachtet werden. Die Bindungen der einzelnen Untereinheiten miteinander ergeben sich durch intermolekulare Wechselwirkungen an spezifischen Stellen der Untereinheitsoberflächen [Voet et al, 2002].

2.3 Faltung und Entfaltung von Proteinen

Proteine werden während der Proteinbiosynthese gebildet. Hierbei wird die DNA mittels der Transkription in mRNA überführt und die gebildete mRNA mittels der Ribosome translatiert. Das Ergebnis der Transkription ist zu Beginn eine einfache Polypeptidkette ohne jegliche Strukturelemente, die Primärstruktur. Dies hat zur Folge, dass in diesem Zustand, dem entfalteten Zustand, das Protein noch keine Funktion besitzt. Um seine Funktion zu erlangen, muss sich das Protein falten, es muss eine räumliche Struktur und seine native Form erlangen, die Tertiär- bzw. Quartärstruktur.

Der Prozess der Faltung beschreibt dabei einen thermischen Vorgang, bei welchem sich das Protein aufgrund der Umgebungsbedingungen und den physiko-chemischen Eigenschaften der einzelnen Aminosäuren (Abbildung 5) innerhalb von Sekundenbruchteilen in die energetisch günstigste Struktur faltet. Bei der Faltung wirken Kräfte auf molekularer Ebene. Dies sind Disulfidbrücken, Van-der-Waals-Kräfte, Salzbrücken, die Polarität sowie die Hydrophobizität der Aminosäuren und Wasserstoffbrückenbindungen. Auf energetischer Ebene betrachtet, sind diese Bindungen schwächer als die Peptidbindungen innerhalb der Polypeptidkette, wodurch sie leichter gebildet und aufgelöst werden können [Voet et al, 2002].

Von den beteiligten Kräften stabilisieren Wasserstoffbrückenbindungen die Struktur am stärksten. Grund dafür ist das sie eine sehr wichtige Rolle bei der Bildung der Sekundärstrukturelemente besitzen. Salzbrücken werden meist an der Oberfläche von Proteinen gebildet und stellen die Verbindungen zwischen Proteinen dar, was zur Ausbildung der Quartärstruktur führt, bzw. anderen Molekülen.

Die Art des Mediums bestimmt, welcher Teil des Proteins dem Medium zugewandt ist. Im Cytoplasma ist der hydrophile Bereich dem Medium zugewandt, während innerhalb einer Membran, einer Lipiddoppelschicht der hydrophobe bzw. der lipophile Teil zugewandt ist. Der dem Medium nicht zugewandte Teil, wird vom Medium abgeschirmt. Durch diese Abschirmung,

den hydrophoben Effekt, wird das Protein dicht zusammen gepackt. [Jaenicke, 1984].

Es ist jedoch noch nicht vollkommen klar, warum sich Proteine so falten, wie sie sich falten. Dieses Faltungsproblem wird mit dem Levinthal-Paradox beschrieben. Bei dem Levinthal-Paradox handelt es sich um ein Gedankenexperiment, bei dem davon ausgegangen wird, dass ein Protein aus 100 Aminosäuren besteht und jede Aminosäure dabei die Möglichkeit besitzt drei Zustände durch zufälliges falten einzunehmen. Dies ergibt also $3^{100} = 10^{48}$ mögliche Strukturen. Desweiteren geht man in diesem Experiment davon aus, dass das Protein für den Wechsel von einem Zustand in den nächsten, eine Nanosekunde braucht. Bei der hohen Anzahl an Möglichkeiten, würde die Faltung mehr Zeit in Anspruch nehmen, als das Universum alt ist [Dressel, 2008]. Daraus folgt, dass die Faltung kein zufälliger Prozess ist, sondern ein nach bestimmten Mustern gerichteter Prozess ist.

Im Prozess der Faltung erlangt ein Protein seine native Form, mittels des Prozesses der Entfaltung kann es wieder in seinen denaturierten Zustand übergehen. Die Entfaltung ergibt sich dabei durch die Änderung der Umgebungsbedingungen. Durch das Einbringen von Tensiden, einer erhöhten Temperatur oder auch eine Änderung des pH-Wertes führen dazu, dass sich die Nebervalenzbindungen des Proteins auflösen und es somit seine räumliche Struktur verliert. Es folgt der Rückfall in die Primärstruktur. Die Entfaltung läuft dabei ähnlich schnell wie die Faltung ab, was dazu führt dass dieser Prozess auch noch nicht nachvollziehbar ist.

Die Entfaltung durch eine der genannten Möglichkeiten hat nicht die Folge, dass das Protein nutzlos wird, denn durch die Wiederherstellung der Ausgangssituation kann das Protein wieder renaturieren und seine native Form wieder herstellen. Die Renaturierung erfolgt dabei bei kleineren Proteinen ohne Funktionsverlust. Bei Proteinkomplexen hingegen, kann es jedoch zu Fehlfaltungen kommen. Dies entsteht dadurch, dass die Faltung der Untereinheiten mit der Assoziation der Untereinheiten miteinander konkurriert. Dabei besitzt die Assoziation einen größeren Effekt als die Faltung, weshalb die

Funktionsfähigkeit des renaturierten Proteins nicht garantiert werden kann [Jaenicke, 1984].

Manche Proteine falten und entfalten sich ständig, da dies durch ihre Funktion notwendig ist. Beispiel dafür ist das Transportprotein Kinesin, welches sich entlang von Mikrotubuli bewegt [Dressel, 2008].

Zur Nachvollziehung der Faltung nutzt man mehrere Faltungsmodelle. Experimente zeigen, dass nicht alle Proteine derselben Faltungsvorstellung folgen. Als Beispiele sollen dabei nur das Two-State-Model und das Molten Globule-Model beschrieben werden.

Bei dem Two-State-Model wird davon ausgegangen, dass sich das Protein nur in seiner denaturierten oder der nativen Struktur vorliegt, dargestellt in Abbildung 9. Die denaturierte Form ist hierbei nicht direkt festgelegt. Es zählen zum denaturierten Zustand alle intermediären Zustände. Die Faltung vollzieht sich hierbei durch die Bildung immer komplexerer intermediärer Zustandsformen bis letztlich der native Zustand erreicht ist. Innerhalb des nativen Zustands besitzen die Aminosäuren den geringsten möglichen Abstand zueinander und dadurch auch die stärksten intramolekularen Kräfte [Zwanzig, 1997].

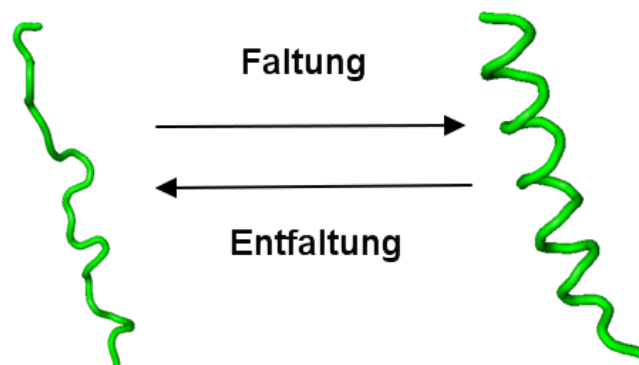


Abbildung 9: Two-State-Model, Protein existiert nur in entfaltetem oder gefaltetem Zustand

Das Molten Globule-Model geht im Gegensatz zum Two-State-Model davon aus, dass das Protein drei Zustände während der Faltung einnehmen kann. Diese Zustände sind der denaturierte, der native und der Molten Globule Zustand. Faltungsbeginn und -ende sind wie bei dem Two-State-Model. Jedoch

gibt es zwischen diesen beiden einen definierbaren Zwischenzustand, den Molten Globule Zustand. Dieser Zustand definiert sich durch 2 grundlegende Eigenschaften. Zum einen besitzt dieser Zustand eine nicht starre Struktur und zum anderen ist die Struktur relativ kompakt. Dieser Zwischenzustand kann experimentell mittels Herabsetzung des pH-Wertes erreicht werden, bei dem sich alle Aminosäuren im sauren Milieu befinden. Dabei zeigt sich, dass der Molten Globule Zustand stark vom denaturierten Zustand abweicht, jedoch nur gering vom nativen. Dieser Zustand weist sowohl Sekundärstrukturelemente als auch hydrophobe Kerne auf, diese Merkmale sind jedoch nicht vollkommen ausgeprägt, da nicht alle Nebervalenzbindungen wie in der nativen Form existieren. Dieser übergeordnete intermediäre Zustand ist nicht mit jedem intermediären Zustand gleichzusetzen. Um als Molten Globule State anerkannt zu werden, müssen drei Kriterien erfüllt sein. Es muss eine hohe Anzahl an Sekundärstrukturelementen vorliegen. Die existierenden Sekundärstrukturelemente, welche keine Beteiligung an der Tertiärstruktur besitzen, erfüllen geometrische Einschränkungen des nativen Zustandes. Die Struktur muss einen hydrophoben Kern besitzen [Haynie & Freire, 1993].

Neben der thermischen Faltung, gibt es auch die kinetische Faltung. Bei dieser Art der Faltung wird nicht die energetisch günstigste Form als nativer Zustand angesehen, sondern das durch den Selektionsdruck geschaffene metastabile lokale Minimum. Der Selektionsdruck wirkt dabei in die Richtung einer bestimmten Funktion. Diese Faltung vollzieht sich meist durch die Hilfe anderer Proteine, wie beispielsweise Chaperone.

Die Unterscheidung zwischen thermischen und kinetischen Falten lässt sich mittels der Energielandschaft der Proteine beschreiben [Dressel, 2008].

3 Energieprofile und Energielandschaften

3.1 Aussagen von Energieprofilen und Energielandschaften

Mittels Energieprofilen und Energielandschaften werden Proteine auf der energetischen Ebene betrachtet. Dabei wird die freie Energie betrachtet, welche sich aus der Summe der Wechselwirkungen zwischen den Aminosäuren und der Umgebung des Proteins ergibt [Heinke, 2010]. Je mehr Wechselwirkungen auftreten, desto geringer ist die freie Energie und umso stabiler das Protein. Dadurch ergibt sich, dass die native Struktur die geringste und der entfaltete Zustand die höchste freie Energie besitzt. Durch den Prozess der Faltung springt dabei die Energie nicht vom Maximum zum globalen Minimum, sondern es verläuft über lokale Minima hin zum globalen Minimum. Hierbei steht jedes lokale Minimum für eine intermediäre Form, welche das Protein während der Faltung einnimmt [Onuchic et al, 1997]. Die einzelnen lokalen Minima werden dabei durch Energiebarrieren getrennt, diese Barrieren zeigen sich durch einen energetischen Anstieg zwischen den Minima [Ivarsson et al, 2008]. Abbildung 10 zeigt das schematische Bild einer Energielandschaft. Durch eine steigende Anzahl an intermediären Zuständen steigt auch der Grad der Frustration eines Proteins. Dies bedeutet, dass mit der steigenden Anzahl an intermediären Strukturen die Höhe der Energiebarrieren zwischen den intermediären Formen sinkt und das Protein zwischen den Zuständen hin und her springen kann, wodurch es schwieriger wird die native Form einzunehmen.

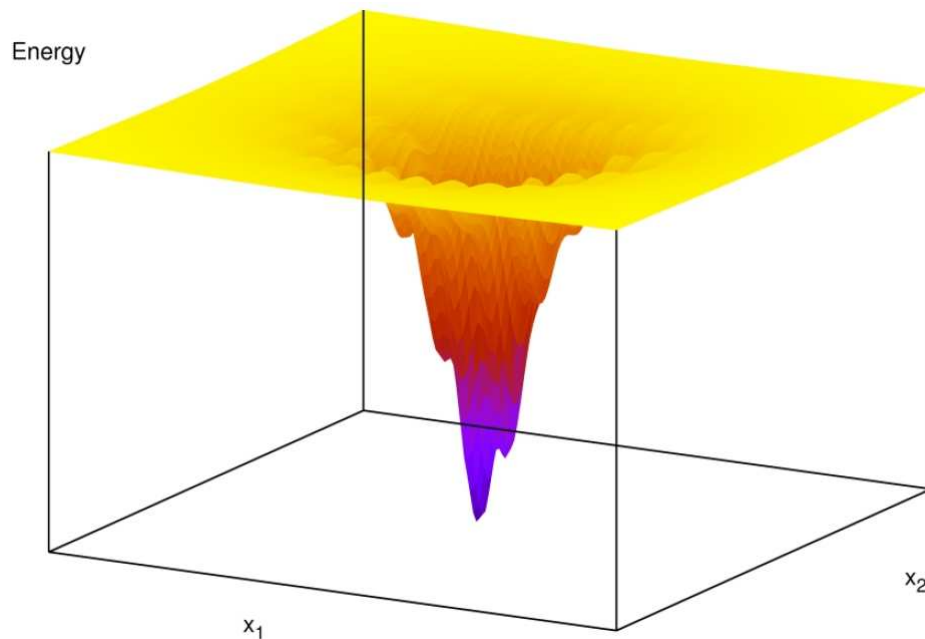


Abbildung 10: Darstellung einer Energielandschaft, zu erkennen ist der Faltungstrichter mit lokalen Minima. x_1 und x_2 stehen für die Koordinaten des Proteins im Zustandsraum [Dressel, 2008].

Mittels Energielandschaften lassen sich zwei Arten von Faltungswegen unterscheiden, der einzigartige und der multiple Faltungsweg. Proteine mit einzigartigen Faltungswegen zeigen, dass sie sich immer auf ein und dieselbe Weise falten. Proteine mit multiplen Faltungswegen, können über mehrere Faltungswege in ihre native Form gebracht werden [Bryngelson et al, 1995].

Neben dem Faltungsverlauf können auch Strukturen mittels Energieprofilen beschrieben werden. Dabei zeigen schon moderate strukturelle Unterschiede große energetische Unterschiede in einem Energieprofil [Onuchic et al, 1997]. Diese Unterschiede können statistisch verwertet werden, wodurch sich auch Strukturen auch vorhersagen lassen [Bryngelson et al, 1995].

3.2 Energiefunktion

Grundlage für diesen Abschnitt bilden die Literaturen [Dressel, 2008][Heinke, 2010].

Für die Berechnung der Energie gilt die Überlegung, dass die Gesamtenergie eines Proteins die Summe aller durch Wechselwirkungen bedingten Energien jeder einzelnen Aminosäure zusammensetzt. Damit ergibt sich folgende Formel:

$$E_{Ges} = \sum_{\langle i,j \rangle} e_{ij}^* f(r_{ij}) + \sum_i e'_{i0} g(i) \quad (1)$$

Hierbei steht e'_{i0} für die Wechselwirkung einer Aminosäure mit dem umgebenden Lösungsmittel, e_{ij}^* definiert die Wechselwirkung zwischen den Resten i und j .

Die Funktion $f(r_{ij})$ relativiert abhängig vom Abstand zwischen Aminosäure i und j die Wechselwirkungen. Für die Bestimmung des Abstands zwischen den Aminosäuren, werden deren C_β -Atome als räumlicher Bezugspunkt genutzt. Diese sind Bestandteil der Seitenkette, welche relativ stark an das C_α -Atom gebunden sind.

Mittels der Funktion $g(i)$ wird der Zustand der Aminosäure i dargestellt. Als Zustände existieren Innen ($g(i) = 1$) und Außen ($g(i) = 0$). Dafür wird die Hydrophobizität der Aminosäure betrachtet. Es wurde festgelegt, dass hydrophobe Aminosäuren vom Lösungsmittel zum inneren des Proteins drängen. Hydrophile Aminosäuren bilden daraufhin Wasserstoffbrückenbindungen mit deren Umgebung.

Um $g(i)$ zu beschreiben, wurde c als Schwerpunkt des inneren Zentrums festgelegt. Mit $r = 10 \text{ \AA}$ eine Kugel um c gesetzt. Dabei gilt:

$$g(i) = \begin{cases} 1: & |C_\alpha - c| < 5 \vee (C_\alpha - C_\beta)(C_\alpha - c) < 0 \\ 0: & \text{sonst} \end{cases} \quad (2)$$

Mittels eines Sets von ca. 2500 globulären Proteinen erstellte man eine Innen/Außen-Häufigkeitsverteilung, siehe Tabelle 2.

Tabelle 2: Häufigkeitstabelle mit Innen/Außen-Zuständen der einzelnen Aminosäure für globuläre Aminosäuren [Dressel, 2008]

Aminosäure	innen	außen	Aminosäure	innen	außen
Cys	4582	1016	His	6419	3366
Ile	20370	4141	Gly	16698	14326
Ser	12576	10411	Asp	10001	14327
Gln	7373	7752	Leu	30615	7107
Lys	9285	15193	Arg	11327	10441
Asn	8225	8928	Trp	4001	1193
Pro	9135	9423	Val	23562	6551
Thr	12537	9622	Glu	11165	18091
Phe	13353	2813	Tyr	11228	3529
Ala	22725	11052	Met	7003	1723

Mittels der Häufigkeitsverteilung kann man durch die Anwendung der Boltzmannverteilung die Energiedifferenz zwischen dem innen und außen Zustand berechnen und somit e'_{i0} .

$$e'_{i0} = -k_B T \ln \left(\frac{n_{in}}{n_{out}} \right) \quad (3)$$

Damit das reine Paarpotential berechnet werden kann, muss (3) umgeformt werden:

$$e_{io} = \left(\frac{1}{i} \alpha_i \right) e'_{i0} \quad (4)$$

Wobei α_i die Anzahl j der Aminosäuren ist, welche in Kontakt mit Aminosäure i stehen. Dabei gilt für einen Kontakt, dass $r_{ij} < 8 \text{ \AA}$ ist.

Für die Berechnung der Wechselwirkung der Aminosäuren untereinander wird dabei folgende Formel verwendet:

$$e_{ij}^* = -k_B T \ln \left(\frac{n_{ij}}{N_{contact} p_i p_j} \right) \quad (5)$$

Hierbei wird die Anzahl der beobachteten Kontakte zwischen den Resten i und j mit n_{ij} beschrieben. $N_{contact}$ gibt die Anzahl aller Kontakte zwischen den Aminosäuren innerhalb der verwendeten 2500 Proteine an. Mit p_i und p_j , siehe Tabelle 3, wird die Wahrscheinlichkeit dafür wiedergegeben, dass eine zufällig gewählte Aminosäure eines Protein der Aminosäure i bzw. j entspricht.

Tabelle 3: Wahrscheinlichkeit p_i dafür das zufällig gewählte Aminosäure gleich der Aminosäure i ist [Dressel, 2008].

Aminosäure	p_i	Aminosäure	p_i	Aminosäure	p_i
Cys	1.8%	Pro	4.5%	Leu	8.9%
Gln	4.0%	Arg	5.1%	Phe	4.1%
Asp	5.8%	His	2.4%	Trp	1.4%
Ser	6.3%	Thr	5.5%	Glu	6.8%
Asn	4.4%	Ala	7.5%	Tyr	3.5%
Val	6.9%	Gly	7.0%	Met	2.0%
Lys	6.3%	Ile	5.7%		

Somit gilt für das einfache Paarpotential:

$$e_{ij} = e_{i0} + e_{j0} + e_{ij}^* \quad (6)$$

Für vorliegende Arbeit wurde eine Änderung bezüglich e_{Ges} vorgenommen, da mittels der Energie Strukturen modelliert werden. Um zu verhindern, dass während der Modellierung eine Struktur gebildet wird, welche nicht den nötigen Abstand zwischen 2 Resten einhält, wurde ein weiteres Kriterium für den Abstand eingeführt, welches die Energie einer Aminosäure auf ∞ setzt.

$$E_{sterisch}(i,j) = \begin{cases} \infty: & r_{ij} \leq 3 \\ 0: & sonst \end{cases} \quad (7)$$

Der Wert aus (7) wird der E_{Ges} hinzugefügt.

$$E_{Ges} = \sum_{\langle i,j \rangle} e_{ij}^* f(r_{ij}) + \sum_i e'_{i0} g(i)e + E_{sterisch}(i,j) \quad (8)$$

4 Einzelmolekülkraftspektroskopie

Einzelmolekülkraftspektroskopie, im englischen Single Molecule Force Spectroscopy (SMFS) ist ein Teilgebiet der Einzelmolekül-Untersuchungsmethoden, engl. Single Molecule (SM) Methods. Diese Methoden betrachten dabei, wie der Name schon sagt, einzelne Moleküle um deren komplexes Verhalten zu erklären. Sie werden neben der Biotechnologie auch in der Chemie, und Physik verwendet. Es gibt viele SM-Methoden, wobei sie nie das gesamte Verhalten erklären, sondern nur spezifische Teile des komplexen Verhaltens von Molekülen. Einen Überblick über die verschiedenen SM-Methoden bietet Abbildung 11. Zu den untersuchten Molekülen zählen hierbei DNA, RNA und Proteine, sowie intra- und intermolekulare Bindungen [Deniz et al, 2008].

		technique	experimental observable	resolution of experimental observable	time resolution
structure	(local) orientation	polarization	polarization or anisotropy	> 5	ms
	short distance	quenching, ET, optical switch	intensity lifetime	< 30 Å	ms
	long distance	magnetic tweezer	force	> sub pN	ms
molecular forces/potential		optical tweezer	force	> sub pN	µs
		AFM	force	> pN	ms
		FRET	intensity lifetime	30 – 100 Å	ms
binding and assembly		FCS	correlation function	ns	ns
		coincidence	coincidence	---	ms
position/movement		particle tracking	PSF	> 1 nm	> ms

Abbildung 11: Übersicht über Single Molecule-Methoden, dargestellt sind die Untersuchungsgebiete, der verwendete Technik, sowie Grenzen der Technik [Deniz et al, 2008].

SMFS-Experimente zählen zu den Manipulationsmethoden der SM-Experimente. Dabei wird mit einer Kraft oder einem Drehmoment auf das zu untersuchende Molekül gewirkt und die Veränderung des Moleküls analysiert. Für solche Experimente werden die optische Pinzette, die magnetische Pinzette oder das Rasterkraftmikroskop, engl. Atomic Force Microscope (AFM), genutzt. Der Aufbau einer SMFS mittels AFM ist in Abbildung 12 dargestellt. Mittels des AFM wird direkt Kraft auf das Protein ausgeübt. Dafür wird es zwischen Cantilever und der Oberfläche des Objektisches befestigt. Die Bindungsstelle zwischen Objektisch – Protein und Protein – Cantilever kann durch die Schaffung von Disulfidbrücken zwischen diesen Punkten bestimmt werden, wodurch es möglich ist die Kraft aus unterschiedlichen Zugrichtungen zu messen und somit unterschiedliche Entfaltungsraten zu untersuchen. Der Tisch wird anschließend in die Z-Richtung bewegt. Durch diese Bewegung beginnt sich der Cantilever zu biegen, wobei diese Biegung mittels einer Photodiode gemessen wird [Schlierf & Rief, 2009][Deniz et al, 2008].

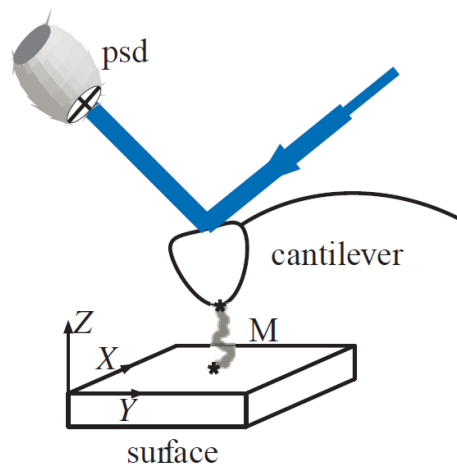


Abbildung 12: Darstellung SMFS mittels AFM, psd - positionspezifischer Detektor, M - Molekül, blau ist der Weg eines Lichtstrahls dargestellt [Deniz et al, 2008].

Durch diese Bewegung werden intramolekulare Bindungen aufgetrennt, was zu einer Instabilität der Sekundärstrukturelemente des Proteins führt und diese dadurch beginnen sich zu entfalten. Es werden nur die Nebenvalenzbindungen, wie Wasserstoffbrückenbindungen getrennt, da diese energieärmer sind als die Hauptvalenzbindungen zwischen den aufeinanderfolgenden Aminosäuren. Die hierbei resultierende Kraft während einer solchen mechanischen Entfaltung nimmt Werte im pN-Bereich, $1 \cdot 10^{-12}$ N, ein [Deniz et al, 2008]. Da die Anzahl von Wasserstoffbrückenbindungen und anderen Nebenvalenzbindungen sich von Struktur zu Struktur unterscheidet, ist die Kraft mit der sich ein Protein und seine Sekundärstrukturelemente entfalten lassen unterschiedlich groß. Es gilt hierbei, je mehr Nebenvalenzbindungen vorhanden sind, desto höher ist die benötigte Kraft für die Entfaltung [Bruno et al, 2005].

Für die Entfaltung von Proteinen gibt es dabei zwei mögliche Methoden, die Force-Clamp und die Velocity-Clamp Methode. Bei der Force-Clamp-Methode wird die Kraft, welche zur Entfaltung führt, konstant gehalten. Dies führt dazu, dass man sowohl das Entfaltungs- als auch das Rückfaltungsverhalten messen kann, wie in Abbildung 13 gezeigt. Diese Methode ist jedoch aufwendig, da durch das Bestreben eine konstante Kraft vorherrschen zu lassen, der Abstand zwischen Cantilever und Objektisch nach jedem Entfaltungsschritt neu justiert werden muss [Janshoff et al, 2000].

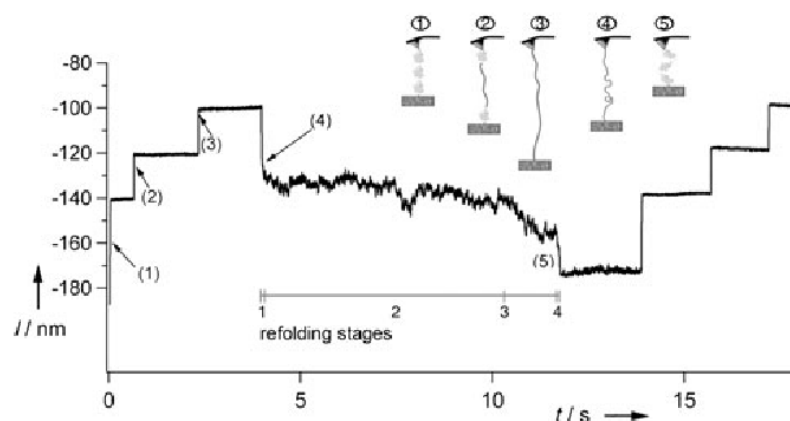


Abbildung 13: Kraft-Abstandskurve mittels Force-Clamp Methode, Entfaltung und Rückfaltung finden nacheinander statt. Der jeweilige Zustand wird mit der Nummer angezeigt [Janshoff et al, 2000].

Bei der Nutzung der Velocity-Clamp Methode wird der Objektisch mit einer konstanten Geschwindigkeit in der Höhe verschoben. Dadurch ergibt sich in der Kraft-Abstandskurve (f-d-Kurve) ein Sägezahnmuster, dargestellt in Abbildung 14. Dabei steht jeder Zahn für die Entfaltung eines bestimmten Teiles des Proteins bzw. für eine Entfaltungsbarriere [Bruno et al, 2005]. Die Höhe des Ausschlages ist dabei von der Entfaltungsgeschwindigkeit abhängig. Je höher die Geschwindigkeit, desto geringer der Kraftausschlag. Durch Rückführung des Objektisches in seine ursprüngliche Position, beginnt das Protein sich wieder in seine native Form zu falten [Janshoff et al, 2000].

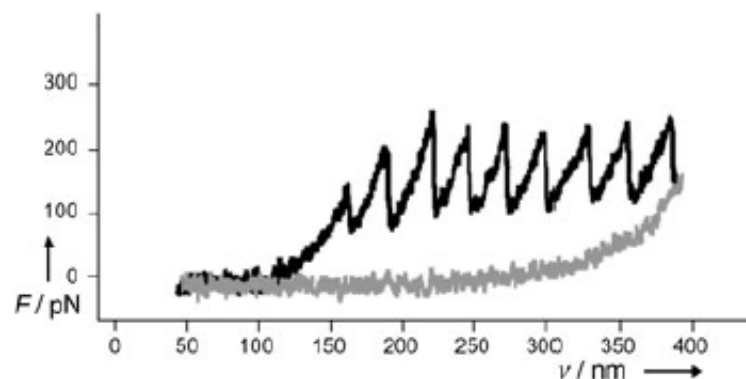


Abbildung 14: Kraft-Abstandskurve mittels Velocity-Clamp Methode. Darstellung eines Multidomänen-Proteins, schwarz - Entfaltung, Grau - Rückfaltung in nativen Zustand [Janshoff et al, 2000].

Bei der Auswertung solcher Experimente ist es wichtig, das umgebende Medium zu beachten, da hierbei große Unterschiede zwischen der Van-der-Waals-Kraft im flüssigen Medium, der Luft oder auch dem Vakuum auftreten. Deshalb sollten bei SMFS-Experimenten immer Bedingungen vorliegen, welche gleich oder ähnlich den in vivo Bedingungen sind [Deniz et al, 2008].

Um von der Kraft und dem End-End-Abstand zwischen Cantilever und Objektisch auf die Anzahl der Aminosäuren zu kommen, welche durch die Kraft entfaltet wurden, gibt es verschiedene Modelle. Hierzu zählen das Freely Jointed Chain Model und das Wormlike Chain Model. Dabei beschreibt das Wormlike Chain (WLC) Model, die Entfaltung des Proteins am besten. Dabei wird das Protein als eine homogene Reihe mit einer Biegeelastizität betrachtet. Die Struktur des Proteins wird somit, wie in Abbildung 15 dargestellt, als eine gebogene, wurmähnliche Kette angesehen. In das Modell fließen dabei sowohl die entropische, als auch die enthalpische Elastizität [Janshoff et al, 2000].



Abbildung 15: schematische Darstellung WLC, l - Länge eines Segmentes und \vec{X} - Vektor zwischen den Enden [Janshoff et al, 2000].

5 Simulation

Für die Simulation der Entfaltung wird ein komplexes Programmpaket bereit gestellt, mit dessen Hilfe die Entfaltung von Proteinabschnitten ermöglicht wird. Theoretische Grundlage für dieses Programmpaket bildet die Dissertationsarbeit von Herrn Frank Dressel [Dressel, 2008]. Mittels dieser Simulation sollen Kraft, End-End-Abstand, sowie die Konformation zu jeglichen Zeitpunkten während des Entfaltens generiert werden. Die Simulation wurde mittels Java umgesetzt und nutzt zur Vereinfachung die BioJava-Bibliothek. Innerhalb des gesamten Programmpaketes werden nur die C_{α} -Atome betrachtet. Eine Übersicht über die erstellten Klassen bietet Abbildung 16.

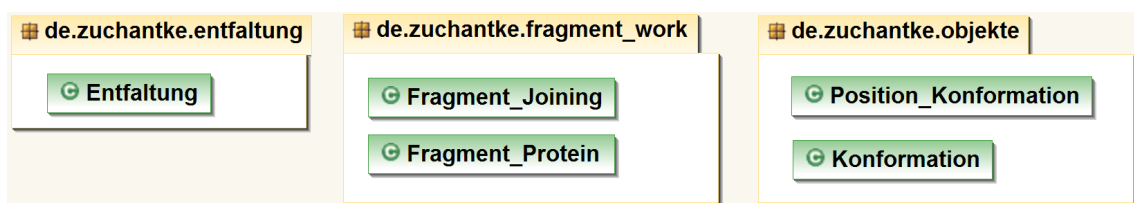


Abbildung 16: Übersicht über erstellte Klassen

5.1 Fragment-Bibliothek

Um die Entfaltung zu simulieren müssen Grundlagen geschaffen werden. So baut die Simulation darauf auf, dass aus einer hohen Anzahl von bereitgestellten Konformationen die günstigste für den jeweiligen Entfaltungsmoment ausgewählt wird. Dafür mussten zuerst einmal diese möglichen Strukturen generiert werden. Dazu wird mittels der statischen Fragmentierung Fragmente der Länge 7 aus einem PDB-SubSet erstellt. Dieses PDB-SubSet beinhaltet über 80 000 Proteine mit einer Ähnlichkeit kleiner gleich 30 % als Auswahlkriterium. Neben der Sequenzähnlichkeit wurden auch Proteine aussortiert, welche mit einer Auflösung größer 2,5 Å aufgenommen wurden. Desweiteren enthielt keine Datei des PDB-SubSets RNA- oder DNA-Fragmente. Da teilweise die Einträge einer PDB-Datei nicht vollständig sind, wurden diese einzelnen PDB-Dateien nochmals auf Fehler überprüft und in ihre jeweiligen einzelnen Ketten aufgeteilt. Dabei erhielt man für die statische Fragmentierung 41 000 fehlerfreie Abschnitte mit einer Sequenzlänge von mindestens 8 Aminosäuren. Über diese Abschnitte wurde

ein Fenster der Größe von 8 Aminosäuren gelegt, um die Fragmente zu generieren. Die Fragmente besaßen jedoch nur die Länge 7, denn die erste Aminosäure diente nur dazu, um die Fragmente später ohne Fehler und mit hoher Korrektheit miteinander zu verbinden. Ein Fragment ist hierbei eindeutig durch die letzte Aminosäure des vorherigen Fragmentes und die ersten beiden Aminosäuren des Fragments beschrieben. Der Aufbau eines Fragmentes ist in Abbildung 17 dargestellt. Es ergaben sich durch die 20 kanonischen Aminosäuren 8002 Fragment-Namen. Dabei besaß jeder Fragment-Name eine sehr hohe Anzahl an Fragmenten. Eine hohe Anzahl an möglichen Fragmenten würde später zu einem erhöhten Rechenaufwand und somit zu einem höheren Zeitaufwand führen.

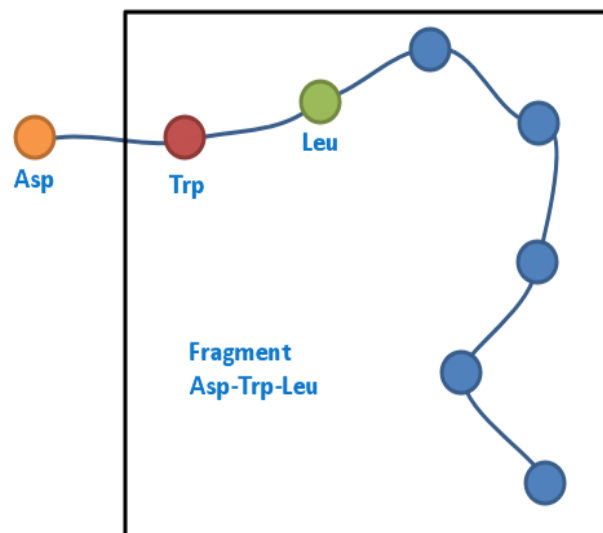


Abbildung 17: Aufbau eines Fragmentes, wobei die Aminosäure außerhalb des Rahmens die letzte Aminosäure des vorhergehenden Fragmentes darstellt und mit den ersten beiden Aminosäuren innerhalb des Rahmens den Namen des Fragments bildet.

Um dem hohen Rechenaufwand entgegen zu wirken, wurden stark ähnliche Fragment-Strukturen zusammengeschlossen. Dies geschah durch ein Clustering der einzelnen Fragment-Namen mittels des FATCAT-Algorithmus [Ye & Godzik, 2003]. Dabei wurde zunächst ein Prototyp für den Strukturvergleich zufällig ausgewählt und gegen alle anderen Strukturen geprüft. Wenn dabei die Wurzel der mittleren quadratischen Abweichung, engl. Root-Mean-Square Deviation (RMSD), kleiner als 0,5 Å war, wurde das

Fragment dem Prototyp zugeordnet. Nach dem alle Strukturen gegen den Prototyp geprüft wurden, wurden die noch nicht zugeordneten gegen alle zu dem Prototyp geordneten Strukturen align und wiederum bei einem kleineren RMSD-Wert als 0,5 Å dem Prototyp zugeordnet. Wenn weiterhin Strukturen existierten, welche einen höheren RMSD-Wert besaßen, wurde ein neuer Prototyp ausgewählt und es wurden wieder die übrigen Strukturen überprüft. Somit erhielt man mehrere Prototyp-Gruppen. Der Ablauf ist in Abbildung 18 dargestellt.

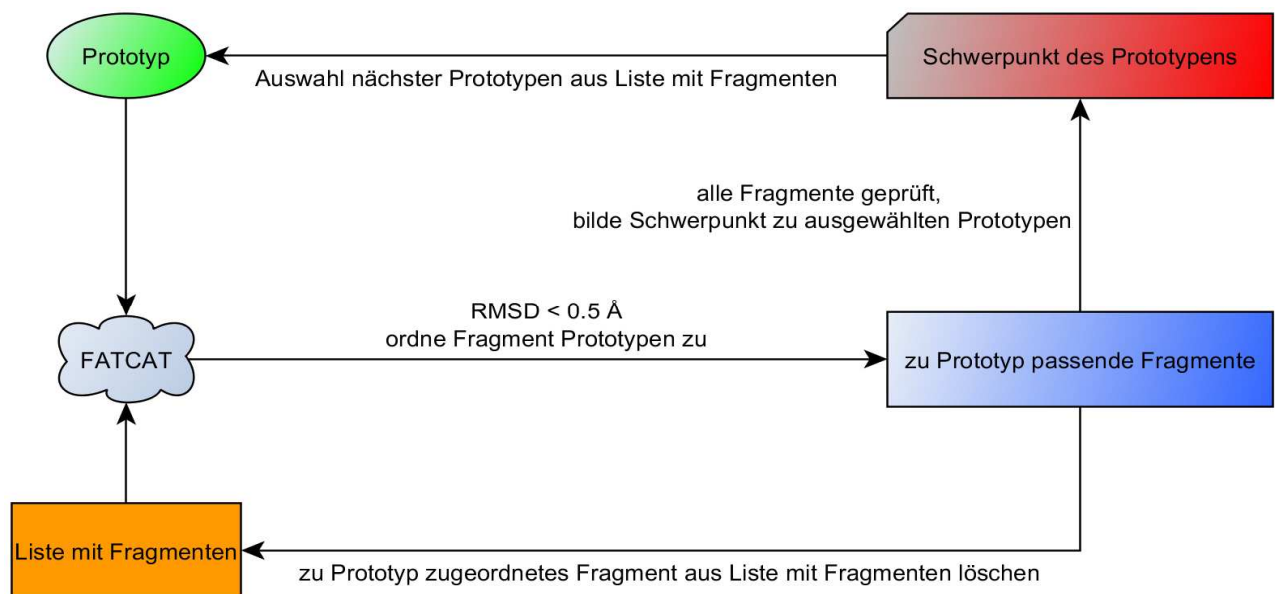


Abbildung 18: schematischer Ablauf Clustering, vereinfachter Darstellung eines Durchlaufes

Aus jeder einzelnen Gruppe wird ein Schwerpunkt erstellt. Dieser besitzt zu jeder Struktur der Gruppe den geringstmöglichen Abstand. Diese Schwerpunkte werden, wenn nötig noch einmal miteinander verglichen und zusammengefasst, sollte ihr RMSD-Wert kleiner 1,2 Å sein.

Durch dieses Clustering ergaben sich durchschnittlich 21 Schwerpunkte pro Fragment-Namen mit einer Standardabweichung von 4,8 Fragmenten. Die Häufigkeitsverteilung ist in Abbildung 19 dargestellt.

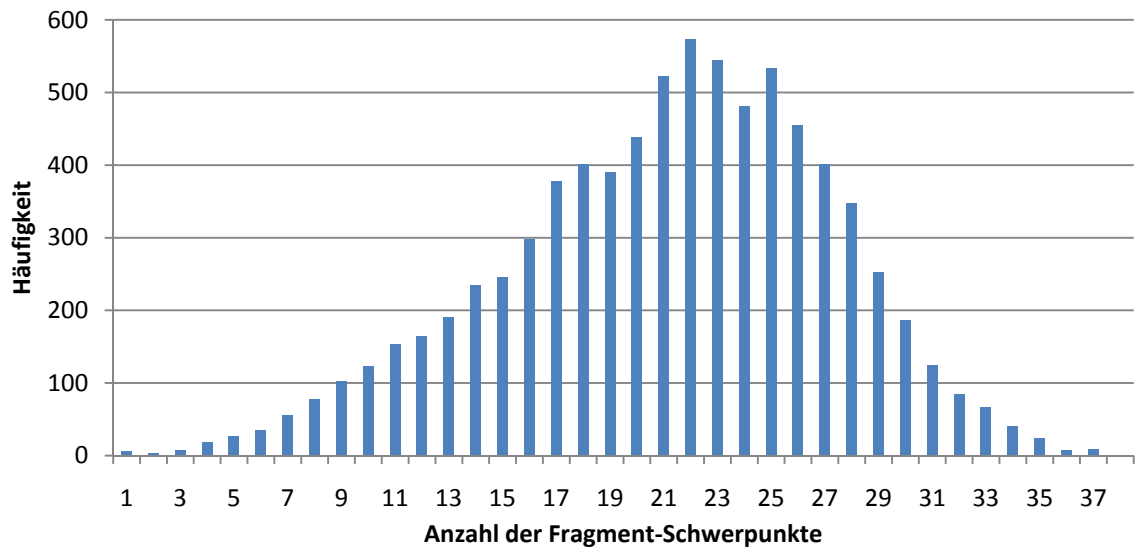


Abbildung 19: Häufigkeitsverteilung der Anzahl an Fragment-Schwerpunkten, über alle Fragment-Namen

Wichtig hierbei, die Fragment-Schwerpunkte stimmen nur in den ersten beiden Resten überein, die restlichen Aminosäuren können von Schwerpunkt zu Schwerpunkt variieren, weshalb sie bei der Modellierung auf das jeweilige Protein angepasst werden müssen.

Mittels der PDB-Datei der betrachteten Struktur werden dessen Fragmente gebildet. Dies geschieht mittels der Klasse *Protein_Fragment*. In dieser wird durch die Methode *aufteilung()* Fragmente der Größe 7 gebildet. Abbildung 20 zeigt die Klasse *Protein_Fragment*. Dabei gibt der Nutzer selbst mit *setStart_rs()* den Rest an, ab welchem Fragmente gebildet werden sollen.

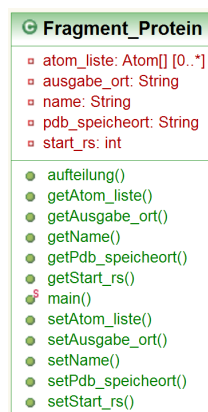


Abbildung 20: Klasse *Protein_Fragment* mit Attributen und Methoden

5.2 Modellierung mittels Branch and Bound

Die erstellten Fragment-Schwerpunktstrukturen der jeweiligen Fragment-Namen bilden die Grundlage für die Modellierung von möglichen Strukturen, welche während der Entfaltung genutzt werden.

Mittels des Branch and Bound Algorithmus werden die Fragmente zuerst testweise miteinander verbunden und bei Gültigkeit der Struktur weiter verfolgt. Dies geschah, bis alle nötigen Fragmente des betrachteten Proteinabschnitts zusammengesetzt wurden. Als Beispiel für die Vorgehensweise soll ein Abschnitt dienen, welcher aus 3 Fragmenten A, B und C besteht. Jedes Fragment besitzt dabei 10 Schwerpunkte. Zuerst wird ein Schwerpunkt aus A ausgewählt. An diesen wird der erste Schwerpunkt aus B angefügt. Sollte diese Teilstruktur gültig sein, so werden alle Schwerpunkte des Fragments C an der Zwischenstruktur angebracht und die Strukturen, welche gültig sind, werden dabei abgespeichert. Sollte jedoch der gewählte Schwerpunkt des Fragments B nicht gültig sein, so wird an dieser Stelle schon der nächste Schwerpunkt des Fragments B gewählt. Dies verkürzt die Rechenzeit um ein vielfaches.

Eine Strukturverknüpfung wird dabei als gültig erachtet, solange sein Energie-Wert unter dem Grenzwert liegt, in diesem Fall unendlich. Der Grund für die Wahl dieses Grenzwertes liegt in Abschnitt 3.2 zugrunde, da dort festgelegt wurde, dass die Energie einer Aminosäure gleich unendlich ist, falls ihr Abstand zu einer anderen Aminosäure kleiner als 3 Å ist. Somit werden nur gültige Strukturen während der Verknüpfung verwendet. Die Energien werden mittels des von Stefan Schildbach erstellten Programms zur Energieberechnung [Schildbach, 2010] bestimmt. Dabei werden alle Proteinabschnitte als globuläre Proteine angesehen.

Als Alternative zur komplexen Strukturberechnungen, bei der Verbindung zweier Abschnitte, werden die Atome eines anzuknüpfenden Fragmentes mittels Koordinatentransformation an das vorherige Fragment angepasst. Dazu wird ein orthogonales Koordinatensystem mittels der Vektoren der letzten drei Aminosäuren des Fragmentes, an welche das neue Fragment gebunden werden sollte, bestimmt.

Das Koordinatensystem ergibt sich dabei aus:

$$\vec{x} = \frac{\vec{A}_{L-1} - \vec{A}_{L-2}}{|\vec{A}_{L-1} - \vec{A}_{L-2}|} \quad (9)$$

$$\vec{y} = \frac{(\vec{A}_L - \vec{A}_{L-1}) - \vec{x}(\vec{A}_L - \vec{A}_{L-1})}{|\vec{A}_{L-1} - \vec{A}_{L-2} - \vec{x}(\vec{A}_L - \vec{A}_{L-1})|} \quad (10)$$

$$\vec{z} = \vec{x} \times \vec{y} \quad (11)$$

Dabei steht \vec{A}_L für den Vektor der letzten Aminosäure und \vec{A}_{L-1} für den vorletzten und \vec{A}_{L-2} den drittletzten Vektor.

Um nun die Positionen der anzuhängenden Aminosäuren zu bestimmen, werden die Koordinaten der x, y und z Ebene wie folgt angepasst:

$$a_{i_k} = a_{i_x} * x_k + a_{i_y} * y_k + a_{i_z} * z_k + A_{L_k} \quad (12)$$

Hierbei steht a_i für die anzuhängende Aminosäure i und k gibt an, welche Ebene des jeweiligen Vektors betrachtet wird.

Wichtig dabei ist, dass zuvor alle Fragmente zum Koordinatenursprung ausgerichtet werden und die jeweils ersten Atome den korrekten Abstand zum Ursprung besitzen, um die Bindung zwischen zwei Fragmenten korrekt darzustellen. Dazu wird der erste Rest auf die Koordinaten {2; 2; 2} gesetzt.

Zusätzlich wird zur Optimierung der genutzten Monte-Carlo-Simulation die Anzahl an gültigen Fragmenten dahingehend geändert wird, das Abschnitte welche einen kürzeren End-End-Abstand besitzen als der Grundzustand der betrachteten Struktur, nicht gespeichert werden.

Mittels der Klasse *Fragment_Joining* wird die Modellierung realisiert. Dazu dienen zwei Methoden. Zum einen die Methode *branch_and_bound()*, welche die einzelnen Fragmente miteinander verknüpft und zum anderen die Methode *verbinde_abschnitte()*, mit welcher sich ein statischer Teil mit einem zu entfaltenden Teil verbinden lässt. In Abbildung 21 ist das Klassendiagramm für *Fragment_Joining* dargestellt.

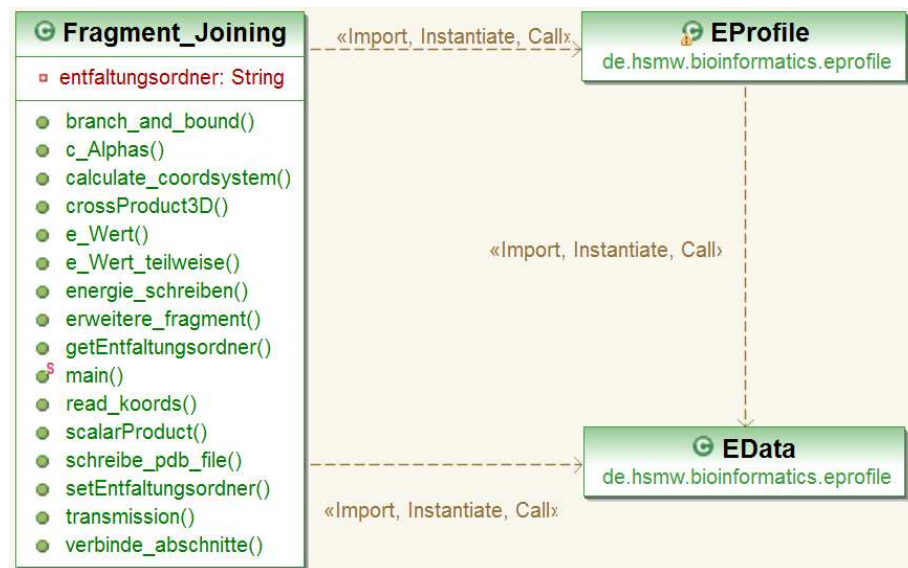


Abbildung 21: Klassendiagramm für das Verbinden von Fragmenten

Für die Verknüpfung der Fragmente zu einer Struktur, mittels der Methode *branch_and_bound()*, werden die darin verbauten Fragmente benötigt. Diese können mittels der Klasse *Fragment_Protein* generiert werden. Danach fällt der Anwender selbst die Entscheidung, welche Fragmente miteinander verbunden werden sollen. Dazu werden die Fragmente der Ausgangsstruktur innerhalb eines Ordners abgespeichert und dessen Name wird an die Klasse *Fragment_Joining* mittels *setEntfaltungsordner()* weiter gegeben. Die Fragment-Schwerpunkte können dann durch die Methode *transmission()* umgeschrieben werden.

Hierbei werden die Aminosäuren in den PDB-Datei der Fragment-Schwerpunkte des angegebenen Fragment-Namens umgeschrieben und zum Koordinatenursprung ausgerichtet. Nach dieser Ausrichtung wird damit begonnen, die einzelnen Fragmente miteinander zu verbinden. Dabei werden drei Methoden verwendet: *erstes_fragment()*, *naechstes_fragment()* und *letztes_fragment()*. Um zu wissen, welches Fragment eingebaut wird, wird jeder Methode ein Index-Wert zugewiesen, welcher dem aktuell verbauten Fragment in der Struktur entspricht.

erstes_fragment() dient dabei der Ausrichtung des ersten Fragmentes, da dies nicht nochmals ausgerichtet werden muss. Das Fragment und seine Energie werden zwischengespeichert. Als Fragment werden dabei alle Fragment-Schwerpunkte des der Fragment-Namens nacheinander getestet. Mit dieser Methode wird die Methode *naechstes_fragment()* aufgerufen.

naechstes_fragment() und *letztes_fragment()* greifen jeweils auf zwischengespeicherte Fragmente zu und bestimmen anhand dieser das Koordinatensystem, mit dessen Hilfe das nächste Fragment angehängt wird.

naechstes_fragment() ist dabei eine rekursive Methode, welche aufeinander folgende Fragmente immer weiter miteinander verbindet und zwischenspeichert, sollten sie unterhalb der Energiegrenze liegen. Dabei werden alle Fragment-Schwerpunkte eines Namens betrachtet. Sollte dabei durch eine sterische Behinderung die Gesamtenergie unendlich betragen, so stoppt die Methode und wählt den nächsten Fragment-Schwerpunkt aus. Wenn dabei jedoch das letzte Fragment verbaut werden soll, so ruft *naechstes_fragment()* die Methode *letztes_Fragment()* auf.

Mittels *letztes_fragment()* werden die Schwerpunkte des letzten Fragment-Namens an die zwischengespeicherte Struktur angehängen. Sollte dabei der Energie-Wert unter dem Grenzwert liegen wird die Konformation mit einer fortlaufenden Nummer gespeichert.

Um die Fragmente mittels einer der 3 Methoden zu verbinden, werden zu Beginn der jeweiligen Methode die C_α -Atome der betrachteten Fragmente mittels der Methode *c_Alphas()* in ein Array eingelesen, wobei die C_α -Atome ihre korrekt fortlaufende Rest-Nummer erhalten. Mithilfe von *read_koords()* werden die Koordinaten dieses Arrays gelesen, wobei hier entweder alle gelesen werden oder nur die letzten drei C_α -Atome für die Bildung eines neuen Koordinatensystems. Das neue Koordinatensystem wird mittels *calculate_koords()* erzeugt. Dabei wird mittels *scalarProduct()* das Skalarprodukt gebildet und mittels *crossProduct3D()* das Kreuzprodukt gebildet. Um nun 2 Fragmente miteinander zu verbinden, werden das erzeugte Koordinatensystem und das Array der C_α -Atome mittels der Methode *erweitere_fragment()* an das vorherige Fragment angepasst.

Für die Bestimmung der Energie der Struktur, gibt es zwei mögliche Methoden. Zum einen *e_Wert()*, mit dem die gesamte Struktur betrachtet wird und *e_Wert_teilweise()*, bei dem der Nutzer die Aminosäure festlegt, ab welcher die Gesamtenergie berechnet werden soll.

Für die Erzeugung der PDB-Dateien der möglichen Konformationen existiert die Methode *schreibe_pdb_file()* und zum Schreiben der Energie-Dateien wird *energie_schreiben()* genutzt.

Trotz der Optimierung der Strukturmodellierung bzw. Strukturgenerierung ist klar, dass durch eine größere Anzahl an zu verbindenden Fragmenten, die Rechenzeit proportional ansteigt.

Bei der Methode *verbinde_abschnitte()* wird der statische Teil angegeben und der Ordner mit den zu verbindenden Konformationen. Daraufhin werden diese beiden Abschnitte miteinander verbunden und deren Energie bestimmt. Sollte die Energie dabei unter dem Grenzwert liegen, so wird die komplette Struktur in einem Ordner gespeichert, welcher gleich dem Namen des Konformations-Ordners ist, zuzüglich des Wortes „komplett“. Als Beispiel soll hier der Konformations-Ordner „Helix C“ genannt werden. Der Ordner in welchem der statische und der jeweilige entfaltbare Teil verbunden sind, heißt „Helix C komplett“.

5.3 Bestimmung einzelner Abschnitte

Um ein Membranprotein mittels der erstellten Simulation zu entfalten, müssen zuerst die Transmembranbereiche (TM-Bereiche) eingeteilt werden. Grund dafür ist die Vorgehensweise des Algorithmus, welcher nur einzelne Strukturen entfaltet und nicht das gesamte Protein betrachtet. Dieser Überlegung ergab sich dadurch, dass zur Ermittlung des Entfaltungsweges die Energie des Proteins betrachtet wird, wie in Abschnitt 3 beschrieben. Wobei hier eine Unterscheidung zwischen globulären Proteinen und Membranproteinen getroffen werden muss. Hierzu nutzt man den von Stefan Schildbach erstellten Algorithmus zur Generierung von Energieprofilen [Schildbach, 2010] zur Bestimmung der Energie einer Struktur. Hierbei wird in einer erweiterten Version die PDB-Datei zu dem Webservices „TMDET“ geschickt, welcher die Transmembranbereiche und Nichttransmembranbereiche (nTM-Bereiche) einer Konformation bestimmt, siehe Tabelle 4. Anhand der Position einer Aminosäure, ob sie sich nun innerhalb der Membran oder außerhalb einer Membran befinden, erhält sie einen spezifischen Energiewert.

Tabelle 4: Ergebnis der TMDET Abfrage zu 1BRR, dabei wurde nur Kette A mit den C α -Atomen betrachtet [URL 4].

Chain 'A' has 7 transmembrane segments:	
Side 1	7-10
Transmembrane	11-26
Side 2	27-45
Transmembrane	46-61
Side 1	62-80
Transmembrane	81-95
Side 2	96-106
Transmembrane	107-123
Side 1	124-139
Transmembrane	140-155
Side 2	156-173
Transmembrane	174-187
Side 1	188-205
Transmembrane	206-220
Side 2	221-232

Die Nutzung des „TMDET“-Service ist aber nur begrenzt für die Simulation zu nutzen, denn bei der Entfaltung wird die Struktur aus der Membran gelöst. Durch das lösen der Aminosäuren aus der Membran erhalten diese einen anderen Energiewert. Dieser Energiewert sinkt mit der Anzahl der aus der Membran gelösten Aminosäuren, was jedoch nicht korrekt ist, da durch die Entfaltung die freie Energie eines Proteins zunimmt und nicht ab. Deshalb kann TMDET in diesem Fall nur dazu genutzt werden, um die TM-Bereiche eines Proteins einzuteilen.

Aufgrund dieser Abweichung der Energiewerte, wird deshalb zur Berechnung der Energie einer Struktur die Methode zur Generierung von Energiewerten globulärer Proteine genutzt, obwohl es sich bei dem Protein um ein Membranprotein handelt.

Um nun einen TM-Bereich zu entfalten, werden der TM-Bereich und der davorstehende nTM-Bereich in der zu entfaltenden Struktur zusammen betrachtet. Das bedeutet beispielsweise für Helix A, dass die Aminosäuren aus Side 1 und der folgenden Transmembrane aus Tabelle 4 für die Entfaltung betrachtet werden. Hierbei ist zu beachten, dass sich die Strukturen aus Fragmenten der Größe von 7 Aminosäuren zusammensetzen, weshalb teilweise auch einige Aminosäuren nach dem TM-Bereich betrachtet werden.

5.4 Entfaltungssimulation

Mittels der gebildeten Fragmente durch den Branch and Bound-Algorithmus, kann die Entfaltungssimulation durchgeführt werden.

Es handelt sich hierbei um eine Monte-Carlo-Simulation, welche aus 1000 einzelnen Entfaltungen besteht. Während der Simulation wird zu jedem Zeitpunkt die gegenwärtige Konformation mit dem dazugehörigen End-End-Abstand, der Energie des Systems und die Auslenkung des Cantilevers bzw. die Kraft welche auf den Cantilever wirkt, berechnet und gespeichert.

Während der Simulation wird die Auslenkung des Cantilevers x zu jedem Zeitpunkt t bestimmt. Der Geschwindigkeit der Bewegung v ist dabei konstant:

$$x = vt \quad (13)$$

Die Auslenkung zu Beginn x_0 , ist gleich dem End-End-Abstand der nativen Form. Mittels der Auslenkung x und der Federkonstante k lässt sich die Federkraft F während des Verbiegens des Cantilevers bestimmen:

$$F = -kx \quad (14)$$

Desweiteren, kann mittels der Auslenkung x und der Federkonstante k die Energie des Cantilevers, E_C bestimmt werden:

$$E_C = \frac{1}{2}kx^2 \quad (15)$$

Aus der Summe der Energie der Konformation im Zustand i , E_i und der Energie des Cantilevers E_C ergibt sich die zeitabhängige Gesamtenergie $E_i(t)$:

$$E_i(t) = E_i + \frac{1}{2}k(Z_P(t) - d_i)^2 \quad (16)$$

Hierbei beschreibt d_i den End-End-Abstand der Konformation im Zustand i .

Zu Beginn befindet sich die betrachtete Struktur in der nativen Form, welche vom Nutzer festzulegen ist. Danach wird angegeben, welche Strukturen genutzt werden sollen, um den Entfaltungsweg der nativen Struktur zu simulieren. Desweiteren gibt der Nutzer die Aminosäure an, ab welcher die Energie der Struktur berechnet werden soll. Der Grund dafür ergibt sich dadurch, dass es auch möglich sein soll, nur einen bestimmten Teil einer definierten Struktur zu entfalten. Da mittels des Branch and Bound Algorithmus auch statische, sich nicht entfaltende Teile an die zu entfaltenden Abschnitte gebracht werden können und die Energie der sich nicht entfaltenden Aminosäuren nicht betrachtet werden sollen.

Zunächst werden durch das Programm die Energie der Struktur, sowie der End-End-Abstand der Struktur bestimmt und $E_i(t)$ der nativen Struktur. Die Zeit t wird um eine Nanosekunde erhöht und es wird zufällig eine Konformation gewählt, welche die nächste im Verlauf der Entfaltung einnehmen soll. Für diese werden alle nötigen Werte bestimmt.

Um nun einen Entfaltungsschritt von Konformation i zu j zu simulieren, wird eine Übergangswahrscheinlichkeit $p_{i \rightarrow j}$ bestimmt:

$$p_{i \rightarrow j} = \min \begin{pmatrix} 1 & : & E_i > E_j \\ e^{-(\Delta E_{ij} - E_C)} & : & E_i \leq E_j \end{pmatrix} \quad (17)$$

Wenn $p_{i \rightarrow j}$ kleiner 1 ist, so wird eine Zufallszahl r zwischen 0 und 1 generiert, wobei sich die aktuelle Konformation E_i wie folgt ergibt:

$$\begin{pmatrix} E_i = E_j : r \geq p_{i \rightarrow j} \\ E_i = E_j : r < p_{i \rightarrow j} \end{pmatrix} \quad (18)$$

Sollte die zufällig gewählte Konformation nicht als nächste Konformation für Entfaltung in Frage kommen, so wird erneut zufällig eine Konformation ausgewählt. Die zufällige Auswahl einer Konformation, wird maximal doppelt solange durchgeführt, wie die Anzahl an möglichen Konformationen ist. Sollte trotz dessen keine Konformation als möglich erachtet werden, so wird die aktuelle Konformation beibehalten, der Zeitschritt erhöht und der nächste Zeitabschnitt betrachtet. Die betrachtete Zeit kann dabei festgelegt werden und wurde im Falle der betrachteten Helices auf 6 ns gesetzt.

Für die Entfaltung einer Struktur ist meist eine sehr hohe Anzahl von möglichen Konformationen vorhanden. Dies hat zur Folge, dass die benötigte Zeit für einen Durchlauf der Simulation sehr hoch ist. Um den Zeitaufwand zu minimieren wird der Algorithmus dahingehend erweitert, dass nach jeder erfolgreichen Findung einer passenden Konformation die Liste der möglichen Strukturen für den nächsten Entfaltungsschritt gekürzt wird. Das Auswahlkriterium ist hierbei der End-End-Abstand der Struktur. Alle Strukturen die einen kleineren Abstand besitzen, werden hierbei für den aktuellen Durchgang nicht mehr beachtet. Um die Auswahl weiterhin zu verkleinern, findet vor Beginn der Simulation eine Gruppierung ähnlicher Konformationen statt. Dabei wird eine zufällige Struktur ausgewählt und zu dieser werden Strukturen zugeordnet, welchen denselben bzw. einen ähnlichen End-End-Abstand besitzen. Von diesen Gruppierungen werden die jeweils 5 energetisch günstigsten für die Entfaltungssimulation verwendet.

Somit ist es durch die Simulation möglich, jegliche intermediäre Form, welche innerhalb einer Entfaltung auftritt, mit der jeweiligen Kraft und Energie darzustellen.

Die Entfaltungssimulation wird mittels der Klasse *Entfaltung*, Abbildung 22, realisiert. Die Klasse besitzt die Methode *start_simulation()*, mit welcher die Simulation gestartet wird. Als Eingaben benötigt sie die Angabe des Ordners, in welchem sich die möglichen Konformationen für die Entfaltung befinden. Desweiteren noch den Namen der Ausgangsstruktur und den Namen unter welchem die Ergebnisse gespeichert werden sollen.

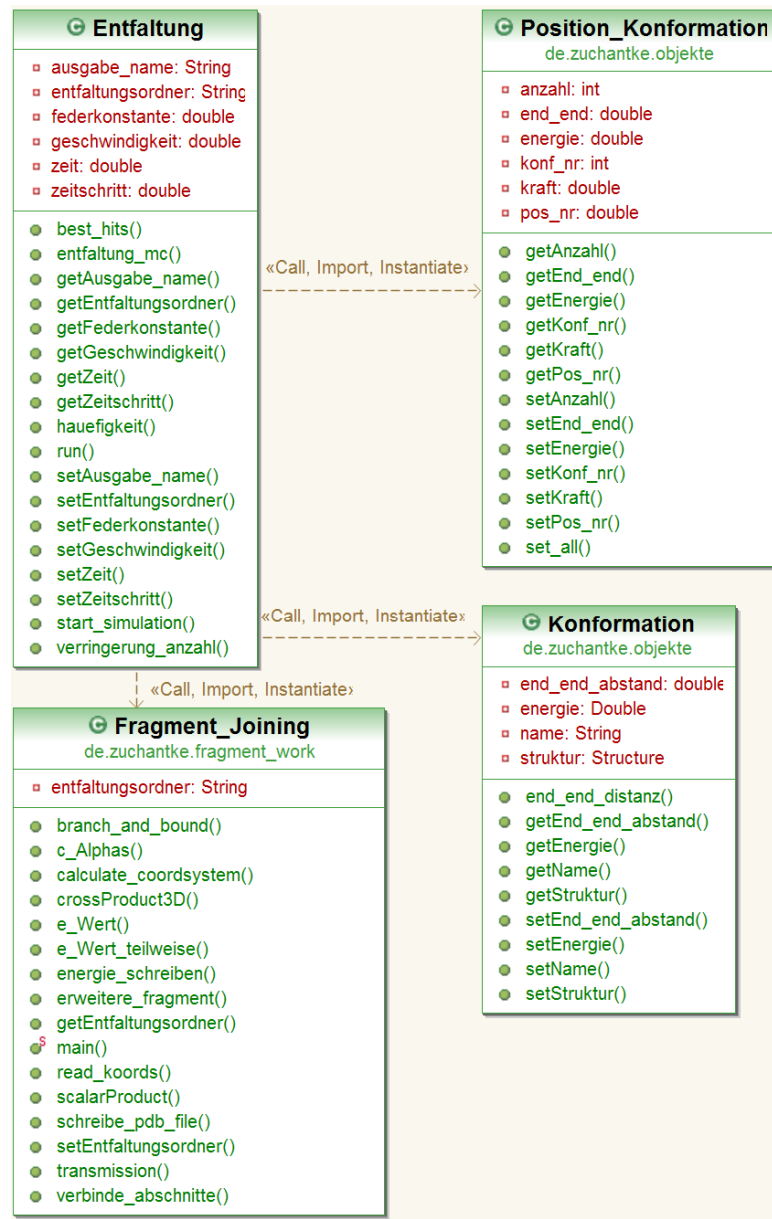


Abbildung 22: Klassendiagramm für die Klasse *Entfaltung*

start_simulation() liest dann mittels der Methode *konformationen()* die einzelnen Konformationen des angegebenen Konformationordners in eine Arrayliste des Typs *Konformation* ein. Dabei werden die für die Entfaltung benötigten Eigenschaften der einzelnen Konformationen berechnet und gespeichert. Die Eigenschaften der Konformationen belaufen sich dabei auf die ID der Struktur, den End-End-Abstand, die Gesamtenergie der Struktur.

Auf Wunsch des Anwenders folgt mittels der Methode *verringderung_anzahl()* die Minimierung der möglichen Konformationen. Diese werden gruppiert und die 5 energetisch günstigsten pro Gruppe für die Entfaltung genutzt. Hierbei gilt ein ähnlicher End-End-Abstand wenn sich zwei Strukturen um maximal 0,01 nm unterscheiden.

Danach wird die angegebene native Struktur mit ihren Eigenschaften eingelesen. Dabei wird die ID für den nativen Zustand auf 1000000 gesetzt. Anschließend wird 3000-mal die Monte Carlo Simulation zur Entfaltung mittels der Methode *entfaltung_mc()* durchgeführt.

Die Methode *entfaltung_mc()* ruft 1000-mal die Methode *entfalten()* auf, welche jeweils eine einzelne mechanische Entfaltung simuliert. Aus den 1000 Ergebnissen wird dann für den jeweiligen Zeitschritt ein Mittelwert mittels der Methode *mittelwert()* gebildet und diese Mittelwerte werden in der angegebenen Ausgabedatei abgespeichert.

Unter dem Mittelwert wird dabei die Konformation pro Zeitschritt angesehen, welche die geringste Abweichung zum Mittelwert des End-End-Abstandes, als auch zum Mittelwert der Energie besitzt.

In der Methode *entfalten()* werden die Eigenschaften der einzelnen Entfaltungsschritte berechnet. Sollte dabei die Kraft zwischen Zeitschritt i und $i + 1$ sinken, so wird die Kraft der verbleibenden Zeitschritte eines Durchlaufes gleich null gesetzt. Ab diesem Zeitpunkt ist der entfaltete Zustand erreicht und die Entfaltung abgeschlossen, weshalb keine weitere Konformation getestet wird. Dies führt zu einer weiteren Verkürzung der Rechendauer pro Durchlauf.

Die gebildeten Mittelwerte werden mittels der Methode *best_hits()* zusammengefasst. Dabei werden für jeden Zeitschritt die drei häufigsten Konformationen bestimmt. Um dabei den Verlauf der Entfaltung einzuhalten, wird geprüft ob die Energie bzw. der End-End-Abstand mit jedem Zeitschritt zu nimmt. Sollte dies nicht der Fall sein, so werden der End-End-Abstand, die Energie und die Kraft der nicht mehr passenden Zeitschritte auf 0 gesetzt.

Um das Auftretens der jeweiligen Konformation pro Zeitschritt zu bestimmen, ist die Methode *hauefigkeit()* zur Bestimmung der Anzahl zu nutzen. Die Mittelwerte der erstellten Ergebnis-Datei der Entfaltungssimulation lassen sich mittels *read()* einlesen.

6. Entfaltung von Bacteriorhodopsin Helix A,B und C

Bacteriorhodopsin, PDB ID 1BRR, ist ein lichtgetriebener Protonentransporter des *Halobacterium salinarum*, einem salzliebenden Bakterium, dargestellt in Abbildung 23 [Dressel, 2008].

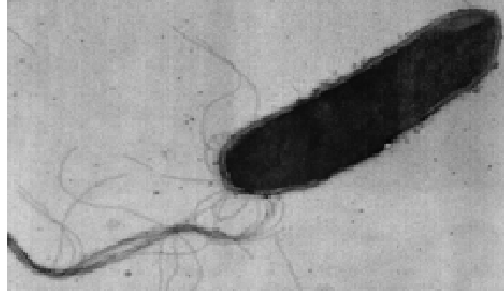


Abbildung 23: *Halobacterium salinarum*, Darstellung ist eine Aufnahme mittels eines Elektronenmikroskops [URL 5]

Mittels Bacteriorhodopsin, dargestellt in Abbildung 24, wird ein Protonengradient erzeugt, welcher dem Bakterium zur Energieerzeugung dient. 1BRR gehört zu den G-Protein-bindenden Rezeptoren [Dressel]. Es zählt mit zu den am ausführlichsten betrachteten Transmembranproteinen [Andreopoulos & Labudde, 2011].

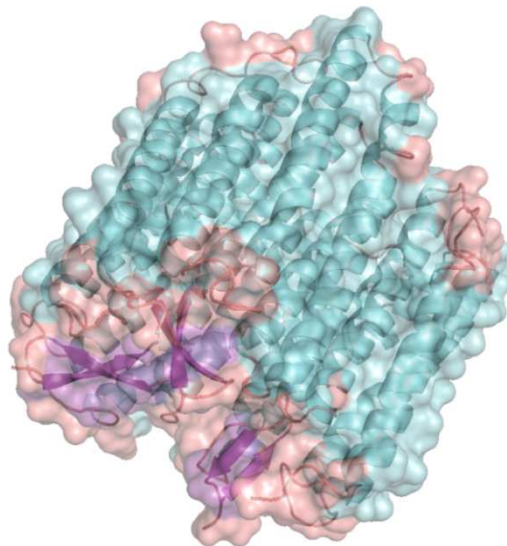


Abbildung 24: Darstellung 1BRR, dabei stellt blau - Helikalebereiche, rot - Coil-Regionen und lila - Faltblätter dar.

Für die Untersuchung des mechanischen Entfaltungsverhaltens bzw. als Test für die erstellte Entfaltungssimulation, wird die Kette A von Bacteriorhodopsin wie in 5.3 beschrieben, nach Tabelle 4 in die jeweiligen Abschnitte für die Helix A, B und C eingeteilt, dargestellt in Abbildung 25. In Experimenten zeigte sich, dass die Geschwindigkeit, mit welcher sich der Objektstisch von der Cantilever entfernt im Fall von 1BRR keinen Einfluss auf das Entfaltungsverhalten hat [Cieplak et al, 2006].

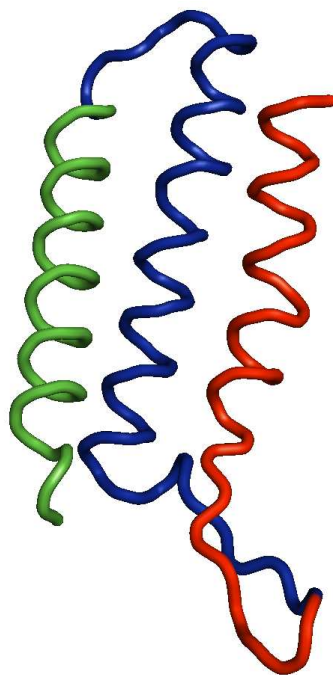


Abbildung 25: Helix A-B-C im nativen Zustand, dabei ist grün Abschnitt Helix A, blau Abschnitt Helix B und rot Abschnitt Helix C dargestellt.

Da in der PDB-Datei von 1BRR nicht alle Atomkoordinaten vermerkt sind reicht diese Kette daher nur von dem Rest 7 bis Rest 232. Durch die Einteilung der Helix-Abschnitte, ergibt sich für A der Sequenzbereich von Aminosäure 8 bis 35, für B 36 bis 70 und für die C 71 bis 98. Die dabei genutzten Fragment-Namen sind in Abbildung 26 dargestellt.

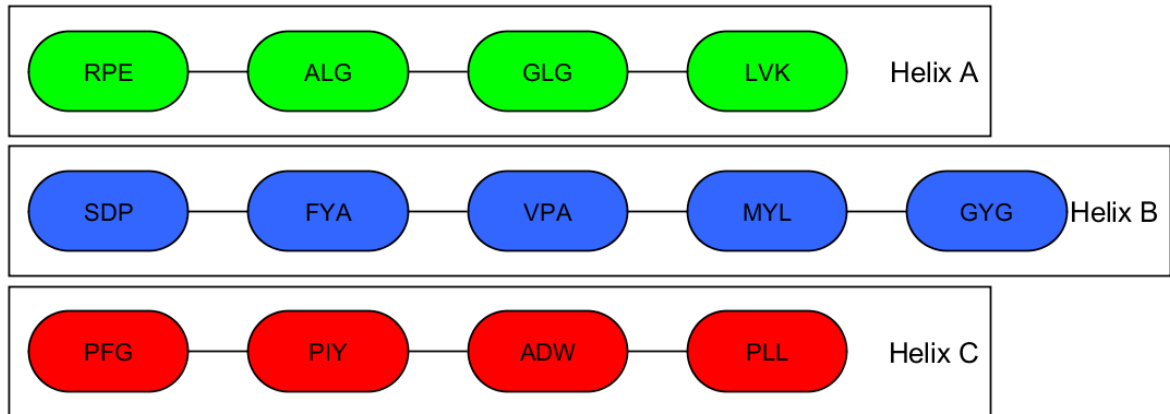


Abbildung 26: schematischer Aufbau der Helices mittels Fragment-Namen. Ein ein Abschnitt besteht dabei wie festgelegt aus 7 Fragmenten

Für die Entfaltung der Helix A, B und C von Bacteriorhodopsin, wurden mittels der in Abschnitt 5.2 vorgestellten *branch_and_bound()* Methode für Helix A und C, jeweils 4 und für Helix B 5 Fragmente miteinander verbunden. Ohne dieses Verfahren hätten 7 062 528 Strukturen für Helix B überprüft werden müssen. Bei Helix A wären es 630 924 und für den Helix C Abschnitt 409 248 Struktur. Durch den Branch and Bound Algorithmus verkürzt sich, durch das Verwerfen von angefangen Strukturen, welche eine sterische Behinderung aufweisen, die Rechenzeit um ein Vielfaches. Es wurden für Helix A 29 716 Strukturen generiert, für Helix B 15 973 und für Helix C 4 489. Die Anzahl an Strukturen ist in Tabelle 5 dargestellt.

Tabelle 5: Übersicht Anzahl gebildeter Strukturen mittels Branch and Bound, dargestellt an Helix A, B und C. Vergleich zwischen mögliche Strukturen, welche durch die verwendeten Fragmente gebildet werden konnten, die durch den Algorithmus tatsächlich gebildeten Strukturen und der Anteil der gebildeten an den möglichen Strukturen.

Helix	Mögliche Strukturen	Gebildete Strukturen	Anteil gebildeter Strukturen
A	630 924	29 716	4,7%
B	7 062 528	15 973	0,2%
C	409 248	4 489	1%

Es zeigt sich somit, dass nur ein sehr geringer Anteil an Strukturen gebildet wird. Der Großteil der Strukturen fällt entweder durch sterische Behinderung oder durch einen zu geringen End-End-Abstand weg, wodurch sich schon eine starke Einschränkung bezüglich möglicher Entfaltungswege zeigt. Jedoch ist die Anzahl an möglichen Strukturen im Fall von Helix A und B sehr hoch und führt zu einer sehr hohen Rechenzeit.

Um die Wechselwirkungen zwischen den benachbarten Helices für die Berechnung der Energie mit einfließen zu lassen, wird vor Helix B und C noch ein statischer Abschnitt vorangestellt. Im Fall des Abschnitts der Helix B ist dies der Abschnitt von Helix A und für den Abschnitt von Helix C wird die Abschnitte von Helix A und B vorangestellt. Durch die statischen Teile senken sich wiederum die Anzahl an verwendbaren Strukturen, wie Tabelle 6 zeigt.

Tabelle 6: Vergleich Anzahl Strukturen in Abhängigkeit vom statischen Teil für die gebildeten Abschnitte der Helix B und C

Helix	Anzahl Strukturen ohne statischen Teil	Anzahl Strukturen mit statischen Teil
B	15 973	8288
C	4 489	3578

Dabei ist zu erkennen, dass für Helix B ca. die Hälfte der möglichen Strukturen wegfällt. Für Helix C sind im Vergleich zu Helix B weniger Strukturen weggefallen.

Dabei wird für eine Verknüpfung zwischen statischen Teil und entfaltbaren Teil nur durch die sterische Behinderung entschieden, ob die Struktur gebildet werden kann. Damit zeigte sich, dass ein Großteil der möglichen Strukturen der Helix B durch ihre Orientierung und ihrer räumlichen Einordnung mit der Helix A kollidieren würden bzw. den selben Raum wie diese einnehmen würde, was nicht möglich ist.

Für die Entfaltungssimulation wurde Anzahl der möglichen Strukturen der drei Helices durch die Methode *verringierung_anzahl()* verringert. In Tabelle 7 ist die Anzahl an möglichen Strukturen mit und ohne die Methode *verringierung_anzahl()* aufgeführt.

Tabelle 7: Vergleich der Anzahl an Konformation in Abhängigkeit der Methode *verringierung_anzahl()* für Helix A, B und C

Helix	Anzahl Konformationen ohne <i>verringierung_anzahl()</i>	Anzahl Konformationen mit <i>verringierung_anzahl()</i>
A	29 716	1830
B	8288	2215
C	3578	1479

Für Helix B und C wurde sowohl eine Entfaltung mit und ohne den statischen Teil durchgeführt. Dies soll den Einfluss der Membran, durch die angefügten statischen Teile simulieren. Hierbei ist davon auszugehen, dass die für die Entfaltung erforderliche Kraft bei den Helices mit statischen Teilen höher ist als bei den Helices ohne. Die Parameter für die Simulation sind in Tabelle 8 aufgelistet.

Tabelle 8: Überblick über verwendete Parameter für Entfaltungssimulation

Parameter	eingesetzte Werte
Entfaltungsdauer	6 Nanosekunden
Entfaltungsschritte	1 Nanosekunde
Federkonstante	0,2 Nm
Zuggeschwindigkeit	654 nm/s

6.1 Ergebnisse Helix A

Die Kraft-Abstandskurve der drei häufigsten Entfaltungsverläufe ist in Abbildung 27 dargestellt. Dabei ist zu erkennen, dass alle Verläufe ein ähnliches Kraftmaxima besitzen, sowie ähnliche End-End-Abstände. Die Steigung der Kraft ist dabei steil, wie in Entfaltungsexperimenten üblich. Dabei zeigt sich ebenfalls, dass die Entfaltung einer Konformation bei annähernd gleichen Kräften auftritt.

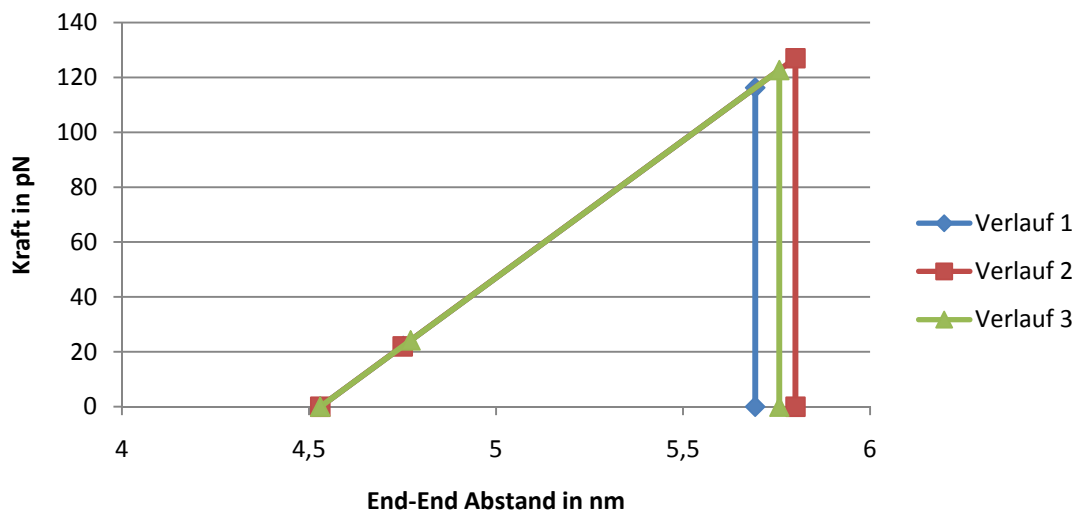


Abbildung 27: Kraft-Abstandskurve der Helix A, aufgeführt sind dabei die 3 häufigsten Verläufe

Betrachtet man den energetischen Verlauf der drei Durchgänge in Abbildung 28, so zeigt sich, dass alle 3 Durchläufe ähnliche energetische Werte besitzen. Dadurch, dass sich alle 3 Verläufe sowohl in den Energien, als auch in der angewendeten Kraft und den End-End-Abständen ähneln, zeigt sich, dass sich die Helix A grundsätzlich auf eine Art und Weise entfaltet. Somit besitzt diese Helix nur einen Entfaltungsweg mit einem intermediären Zustand.

6. Entfaltung von Bacteriorhodopsin Helix A,B und C

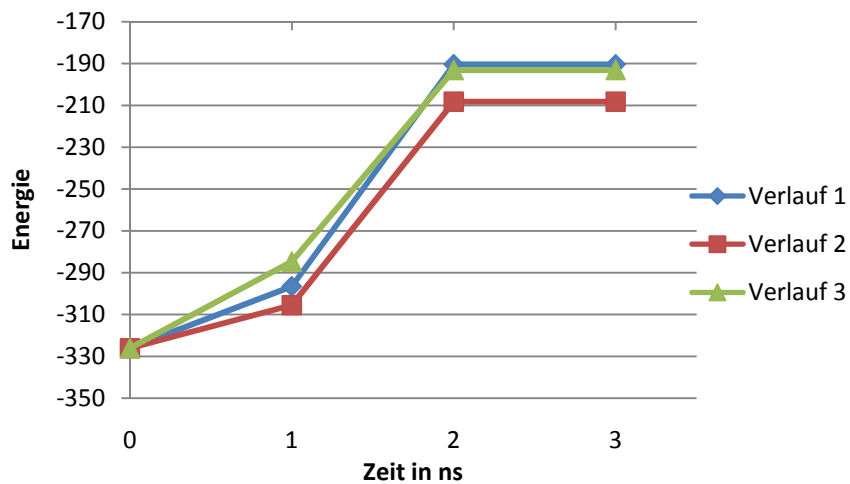


Abbildung 28: Energetischer Entwicklung der Entfaltung von Helix A, dabei zeigen sich nur geringe Unterschiede zwischen den einzelnen Verläufen

Vergleicht man nun die gebildeten Konformationen zu jedem Zeitpunkt der Simulationen in Abbildung 29, so zeigt sich, dass deren Unterschiede nur minimal sind und die Annahme stützen, dass nur ein Entfaltungsweg existiert.

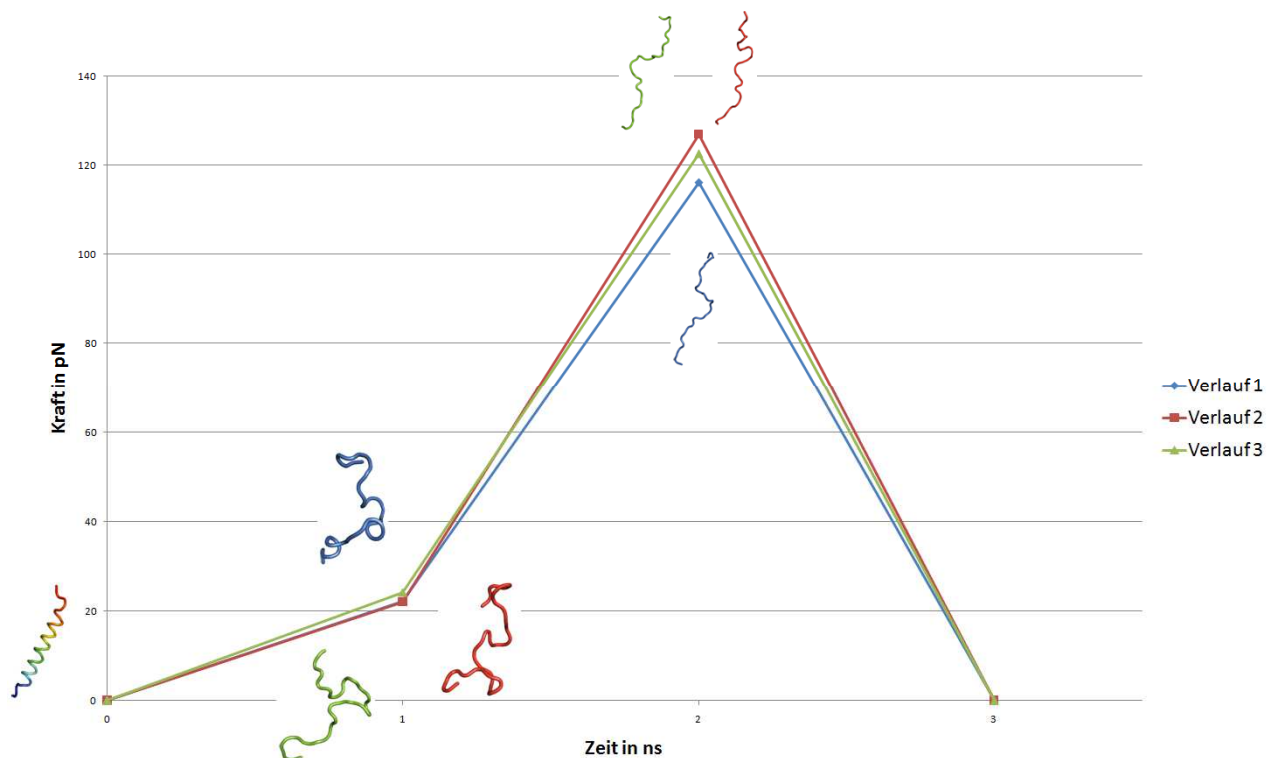


Abbildung 29: Kraft über Entfaltungsdauer der Helix A mit jeweiligen Strukturen, dabei korrespondiert die Strukturen zu den farblich passenden Verläufen.

Um die geringen Unterschiede zwischen den einzelnen Verläufen zu zeigen, wurden deren RMSD-Werte mittels PyMOL bestimmt und sind in Tabelle 9 dargestellt. Dabei wurde jeweils der letzte RMSD-Wert genutzt, welcher PyMOL berechnete. Hierbei gilt für einen RMSD-Wert unter 6,5 Å eine gute Übereinstimmung [Dressel, 2008]. Die Werte zeigen, dass sich die einzelnen Strukturen nicht stark voneinander unterscheiden. Wobei der intermediäre Zustand im Verlauf 3 eine relativ hohe Abweichung aufweist.

Tabelle 9: Vergleich RMSD-Werte für die Konformationen in Helix A, dargestellt sind dabei die beiden gebildeten Konformationen mit den jeweiligen RMSD-Werten pro Zustand zueinander.

	Verlauf 1	Verlauf 2
Intermediärer Zustand		
Verlauf 2	5,281 Å	-
Verlauf 3	6,366 Å	7,534 Å
Entfalteter Zustand		
Verlauf 2	4,745 Å	-
Verlauf 3	0,001 Å	4,782 Å

Diese Abweichung ist aber aufgrund der Häufigkeit des Auftretens der einzelnen Konformation pro Verlauf, aufgeführt in

Tabelle 10, als vernachlässigbar zu betrachten.

Tabelle 10: Häufigkeit der aufgetretenen Konformationen pro Verlauf in Helix A

Zeitschritt in ns	Verlauf 1	Verlauf 2	Verlauf 3
1	0,93 %	0,83 %	0,77 %
2	3,97 %	3,43 %	3,1 %

6.2 Ergebnisse Helix B

6.2.1 Entfaltung mit statischem Teil

Für die Simulation der Entfaltung mit der geminderten Anzahl an Konformationen ergab sich die in Abbildung 30 dargestellte Kraft-Abstandskurve. Dabei zeigt sich wieder ein steiler Kraftanstieg bei einem größer werdenden End-End-Abstand. Dabei weisen alle drei Verläufe wiederum ähnliche Maxima auf. Dabei weisen der Verlauf 1 und 2 jeweils eine intermediäre Konformation, Verlauf 2 hingegen weist zwei intermediäre Formen auf. Dabei zeigt sich das die intermediäre Form von Verlauf 1 eine ähnliche Kraft sowie End-End Abstand, wie die erste intermediäre Form von Verlauf 2 besitzt. Die intermediäre Form von Verlauf 3 liegt dabei ähnlich der zweiten intermediären Form von Verlauf 2. Diese zeigt, dass die Entfaltung der drei Verläufe prinzipiell gleich verläuft und somit ein eindeutiger Entfaltungsweg vorliegt.

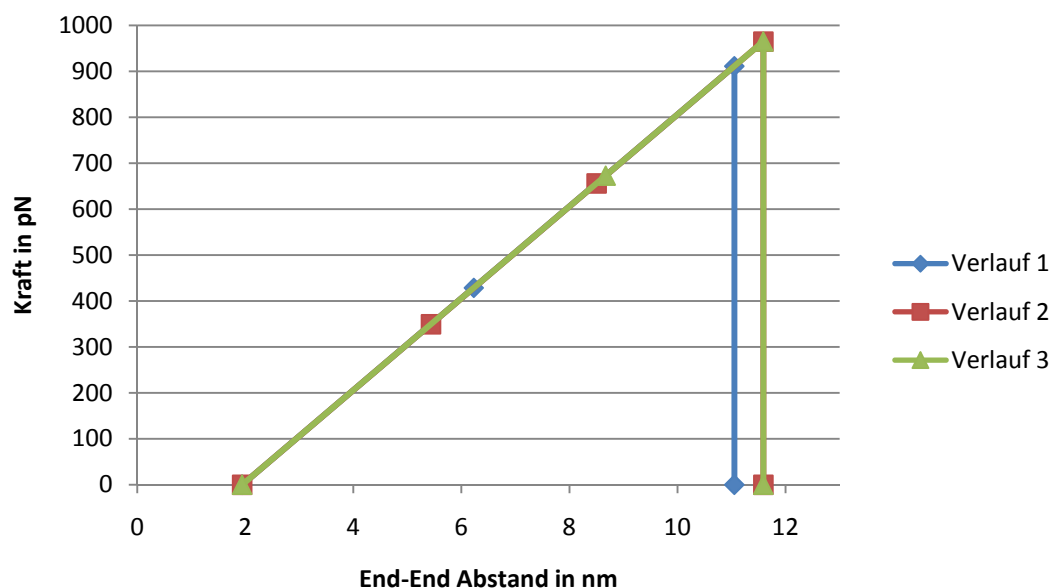


Abbildung 30: Kraft-Abstandskurven für Helix B mit statischem Teil, dargestellt sind die 3 häufigsten Verläufe.

Betrachtet man nun jedoch den Verlauf der Faltung aus energetischer Sicht in Abbildung 31, so zeigt sich ein starker Unterschied zwischen den Energien der Konformationen in der ersten Nanosekunde. Dabei liegen die Energien des Verlaufs 1 und 2 näher beieinander und die Energie der Konformation im Verlauf 3 ist zu diesem Zeitpunkt sehr viel geringer. Ansonsten besitzen alle 3 Verläufe ähnliche energetische Werte. Es zeigt sich bei Verlauf 2, dass dessen Energie nach der zweiten Nanosekunde etwas fällt, jedoch der End-End Abstand steigt. Dies ergibt sich durch die Einbeziehung des statischen Teils.

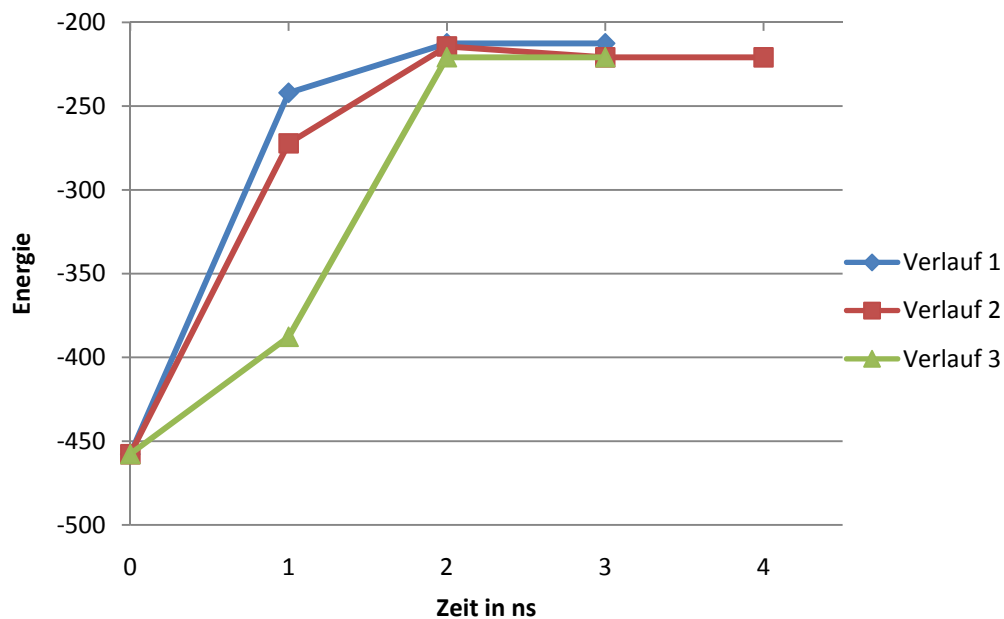


Abbildung 31: Energetische Entwicklung der Entfaltung von Helix B mit statischem Teil, dargestellt sind dabei die 3 häufigsten Verläufe

Mittels der in Abbildung 32 dargestellten Entwicklungen der Kraft über die Zeit für alle drei Verläufe, zeigt sich, dass sich die entfaltenden Zustände in Bezug auf die Kraft sehr stark ähneln und auch die dabei gebildeten Strukturen sehr ähnlich sind. Für Verlauf 2 zeigt sich dabei die zweite intermediäre Form als passende Zwischenform zwischen der ersten intermediären und der entfalteten Form. Desweiteren zeigt sich die vermutete strukturelle Ähnlichkeit der intermediären Form des Verlaufes 1 zu Verlauf 2, sowie die des Verlaufes 3 zu Verlauf 2.

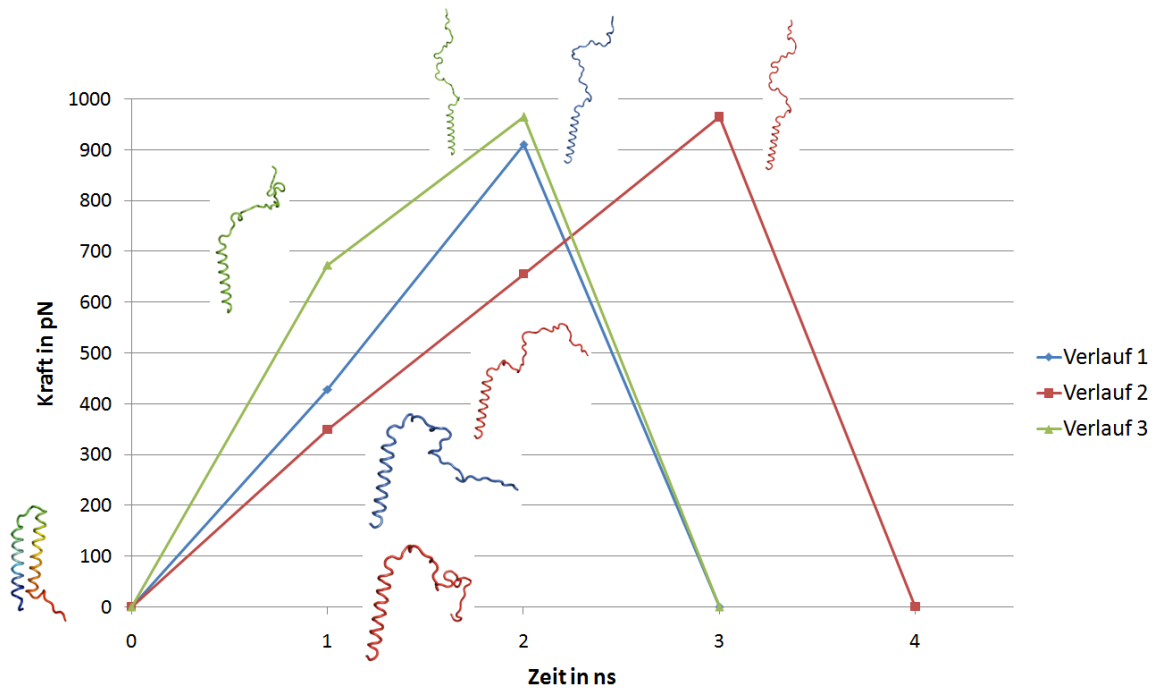


Abbildung 32: Kraftverlauf über die Entfaltungsdauer von Helix B mit statischen Teil, zu dem jeweiligen Verlauf sind dessen Strukturen zum jeweiligen Zeitpunkt dargestellt

Um diese Ähnlichkeiten der Zustände besser zu vergleichen, wurden wiederum die RMSD-Werte der intermediären Formen berechnet. Diese sind in Tabelle 11 aufgelistet, wobei die erste intermediäre Form von Verlauf 2 mit der intermediären von Verlauf 1 und die zweite Zwischenform des Verlaufs 2 mit der Zwischenform des Verlaufs 3 verglichen wird. Dabei zeigt sich, dass sich die jeweiligen Strukturen sehr stark voneinander unterscheiden.

Tabelle 11: RMSD-Werte der intermediären Strukturen der Helix B, dabei wird die erste intermediäre Form von Verlauf 2 mit der intermediären Form von Verlauf 1 verglichen. Die zweite Zwischenstruktur mit der intermediären Struktur von Verlauf 3

Verlauf 2	
Erste intermediäre Konformation	
Verlauf 1	7,289 Å
Zweite intermediäre Konformation	
Verlauf 3	8,886 Å

Um das Entfaltungsverhalten weiterhin zu bewerten, wurde wieder eine Häufigkeitsverteilung der einzelnen Konformation angefertigt. Die Werte sind in Tabelle 12 aufgelistet. Dabei besitzt Verlauf 3 keinen Wert für die dritte Nanosekunde, da dieser Wert angeglichen wurde und die eigentlich darauf folgende Konformation keinen Zusammenhang mehr zeigte.

Tabelle 12: Häufigkeit der Konformationen pro Verlauf für Helix B mit statischen Teil

Zeit in ns	Verlauf 1	Verlauf 2	Verlauf 3
1	0,27 %	0,23 %	0,23 %
2	1,23 %	1,13 %	1,07 %
3	6,5 %	5,1 %	-

Aufgrund des geringen Auftretens der Konformationen in den Verläufen und der stark abweichenden Strukturen lässt sich unter den genutzten Parametern keine eindeutige Aussage über das Entfaltungsverhalten mittels der verwendeten Auswertungsmethode treffen. Durch die Simulation lässt sich nur ein grobes Verhalten abschätzen. Alle drei Verläufe zeigen, dass die Struktur sich sofort von der statisch angefügten Helix A abspreizt und sich dabei schon entfaltet.

6.2.2 Entfaltung Helix B ohne statischen Teil

Ohne den statischen Teil ergibt sich ebenfalls eine steile Kraft-Abstandskurve, wie in Abbildung 33 dargestellt. Jedoch ist dabei die Kraft, welche für die Entfaltung genutzt wird, sehr viel geringer. Neben der Kraft ist auch der End-End Abstand im Vergleich geringer, da der Befestigungspunkt zu Tischoberfläche anders liegt als bei der Helix B mit statischem Teil. Die verringerte Kraft ergibt sich durch das Fehlen des statischen Parts und dessen Einfluss auf die Helix B. Die Kraft ist dabei jedoch im Vergleich zu Helix A höher, was sich durch die höhere Anzahl an Aminosäuren, welche in dem Abschnitt verbaut sind, ergibt. Dabei liegen Maxima von Verlauf 2 und 3 näher

beieinander als Verlauf 1 zu den beiden. Alle drei Verläufe besitzen wiederum jeweils eine intermediäre Form. Dabei unterscheiden sich der End-End Abstand und die angewendete Kraft für die Zwischenformen von Verlauf 1 und 2 nur gering voneinander.

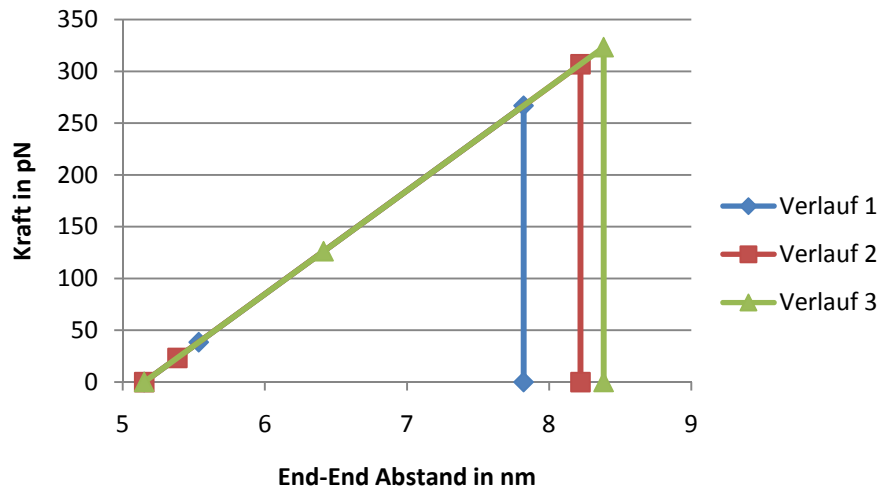


Abbildung 33: Kraft-Abstandskurve für Helix B ohne statischen Teil

Bei der energetischen Entwicklung der Helix B in allen 3 Verläufen, zeigt sich in Abbildung 34, dass diese eine hohe Ähnlichkeiten zueinander besitzen. Was wiederum einen Unterschied zu dem Verlauf mit dem statischen Teil darstellt. Jedoch zeigt sich, dass sie ähnliche Energien besitzen, aber teils größere Abstandsunterschiede aufweisen.

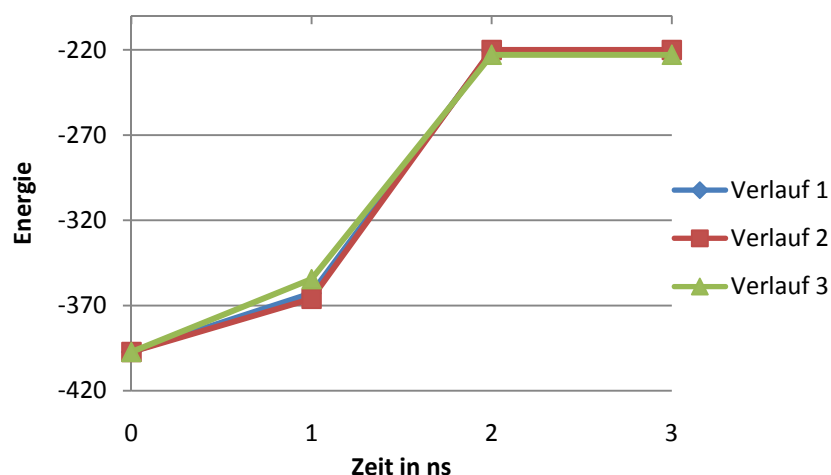


Abbildung 34: Energetische Entwicklung der Entfaltung von Helix B ohne statischen Teil

Diese starken strukturellen Unterschiede zeigen sich allerdings nicht in der Gesamtenergie der Strukturen. Um zu verstehen wie sich die Gesamtenergie bildet, werden die Energieprofile der drei Strukturen in Abbildung 35 betrachtet und zeigen dabei teils große Unterschiede zwischen den Profilen. Dies lässt auf stark unterschiedliche Strukturen hindeuten.

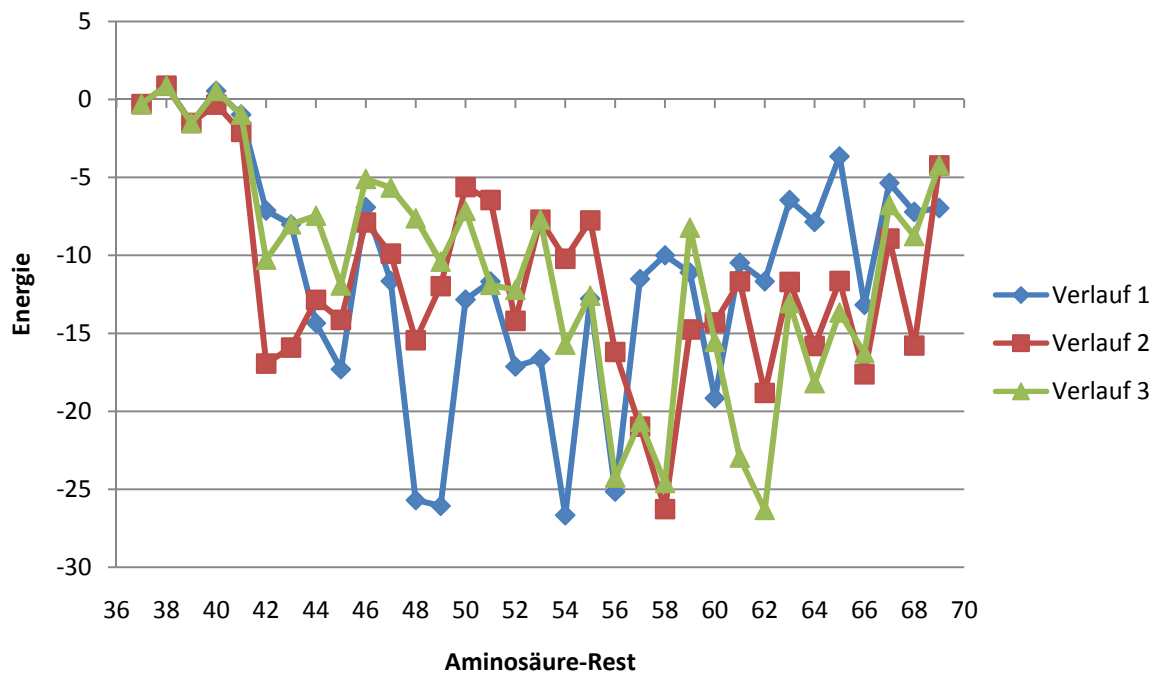


Abbildung 35: Energieprofile der intermediären Form der Helix B ohne statischen Teil. Dabei zeigen sich große energetische Unterschiede zwischen den einzelnen Profilen

Um diese Annahme, über stark unterschiedliche intermediäre Strukturen, zu bestätigen, betrachtet man den Kraftverlauf über die Entfaltungsdauer in Abbildung 36. So zeigt sich deutlich der Unterschied zwischen Verlauf 3 und den anderen beiden Verläufe. Dabei ist der gesamte Kraft beim Verlauf 3 höher als die anderen beiden. Es wird dabei auch mehr Kraft benötigt, um die native Struktur aufzulösen.

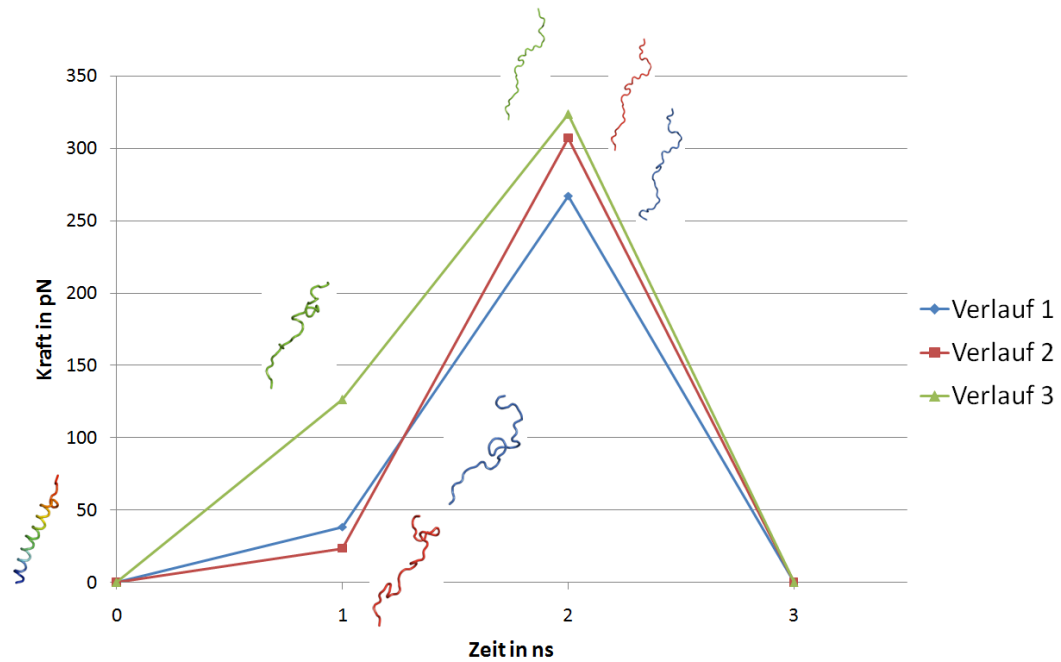


Abbildung 36: Kraftverlauf über die Entfaltungsdauer von Helix B ohne statischen Teil

Für die Verdeutlichung der Unterschiede zwischen den intermediären und entfalteten Formen, wurden dafür die jeweiligen Konformationen miteinander verglichen und deren RMSD-Werte gemessen. Diese Werte sind in Tabelle 13 dargestellt und zeigen, dass sich die entfalteten Strukturen nur vernachlässigbar gering voneinander unterscheiden. Hierbei wäre jedoch mindestens ein Wert von 1,2 Å zu erwarten, da für die Bildung Fragment-Schwerpunkte genutzt wurden, welche sich in mindesten 1,2 Å voneinander unterscheiden. Der Grund dafür liegt in der Vorgehensweise, wie mittels PyMOL der RMSD-Wert gebildet wird. Dabei werden Durchgänge gestartet bei denen teilweise Atome nicht mit beachtet werden und somit flexibler sind. Die intermediären Strukturen weichen dabei mit mehr als 6,5 Å voneinander ab, was zeigt, dass diese sich stark voneinander unterscheiden, wie schon die Energie-Profile vermuten ließen.

Tabelle 13: RMSD-Wert der auftretenden Strukturen bei Helix B ohne statischen Teil. Dabei wurden die intermediären und die entfalteten Strukturen separat verglichen.

	Verlauf 1	Verlauf 2
Intermediäre Struktur		
Verlauf 2	7,097 Å	-
Verlauf 3	7,285 Å	6,979 Å
Entfaltete Struktur		
Verlauf 2	0,002 Å	-
Verlauf 3	0,002 Å	0

Die Häufigkeit der Konformationen ist dabei leicht höher, als bei Helix B mit dem statischen Teil. Sie sind in Tabelle 14 aufgeführt und zeigen, dass alle 3 Verläufe ähnlich oft auftreten.

Tabelle 14: Häufigkeit der Konformationen in allen Verläufen von Helix B ohne statischen Teil

Zeit in ns	Verlauf 1	Verlauf 2	Verlauf 3
1	0,37 %	0,33 %	0,33 %
2	2 %	1,63 %	1,57 %
3	7,93 %	7,5 %	4,7 %

Im Vergleich zeigt Helix B ohne statischen Teil ein anderes Entfaltungsverhalten als Helix B mit dem statischen Teil, was einen starken Einfluss des statischen Teils und der damit simulierten Membran darstellt.

6.3 Ergebnisse Helix C

6.2.1 Entfaltung mit statischem Teil

Bei der Betrachtung der drei Kraft-Abstandskurven von Helix C in Abbildung 37 zeigt sich wiederum ein steiler, typischer Anstieg in allen 3 Verläufen. Dabei weist jeder Verlauf eine einzige intermediäre Form auf. Somit ist zwischen diesen Strukturen ein hoher Unterschied bezüglich der Kraft und End-End-Abstand zu erkennen. Der entfaltete End-End Abstand ist dabei bei allen 3 Verläufen ähnlich groß, mit ca. 8 Nanometern.

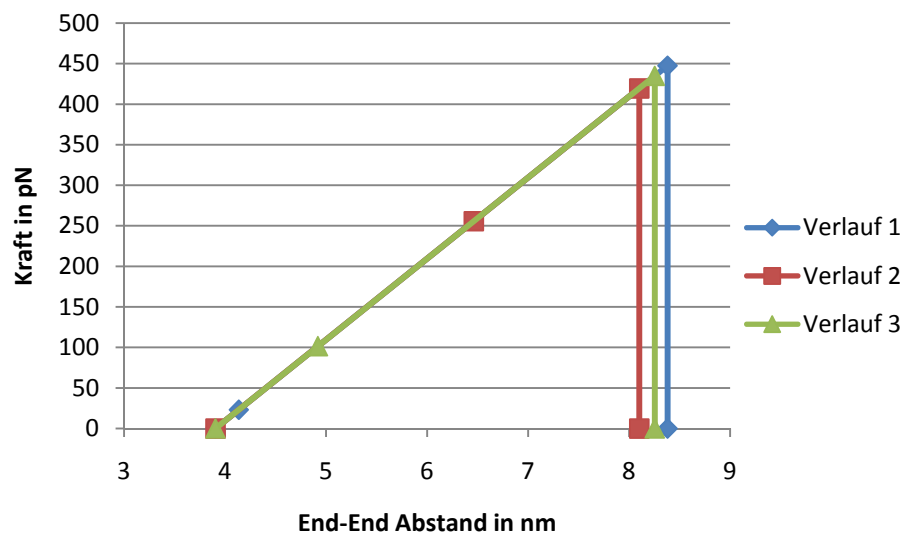


Abbildung 37: Kraft-Abstandskurve Helix C mit statischem Teil

Überraschend bei diesen stark unterschiedlichen Kräften sind jedoch die energetischen Verläufe der drei Entfaltungswege in Abbildung 38. Dabei zeigt sich bei allen drei Verläufen eine sehr ähnliche energetische Entwicklung. Dies lässt darauf schließen, dass sich ähnliche Strukturen bilden. Doch diese energetische Ähnlichkeit steht im Widerspruch mit den End-End Abständen der Zwischenstruktur der 3 Fälle.

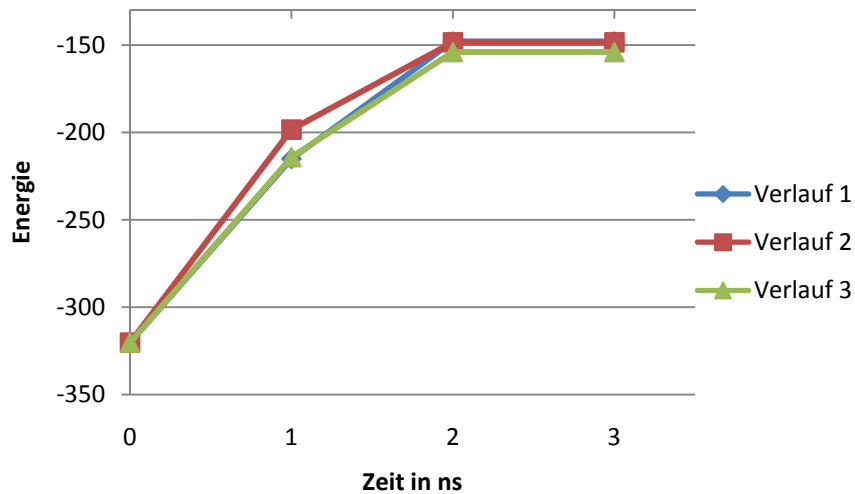


Abbildung 38: Energetischer Entwicklung der Entfaltung von Helix C mit statischem Teil

Um den Widerspruch zwischen den Ähnlichkeiten der Energien und den Unterschieden der End-End Abstände, sowie der Kraft zu verstehen, wurden für die drei intermediären Formen jeweils ein Energieprofil, Abbildung 39, erstellt wurde. Dabei zeigen sich zwischen allen 3 Formen nur geringe Unterschiede. Diese erklären jedoch nicht die großen End-End Abstände.

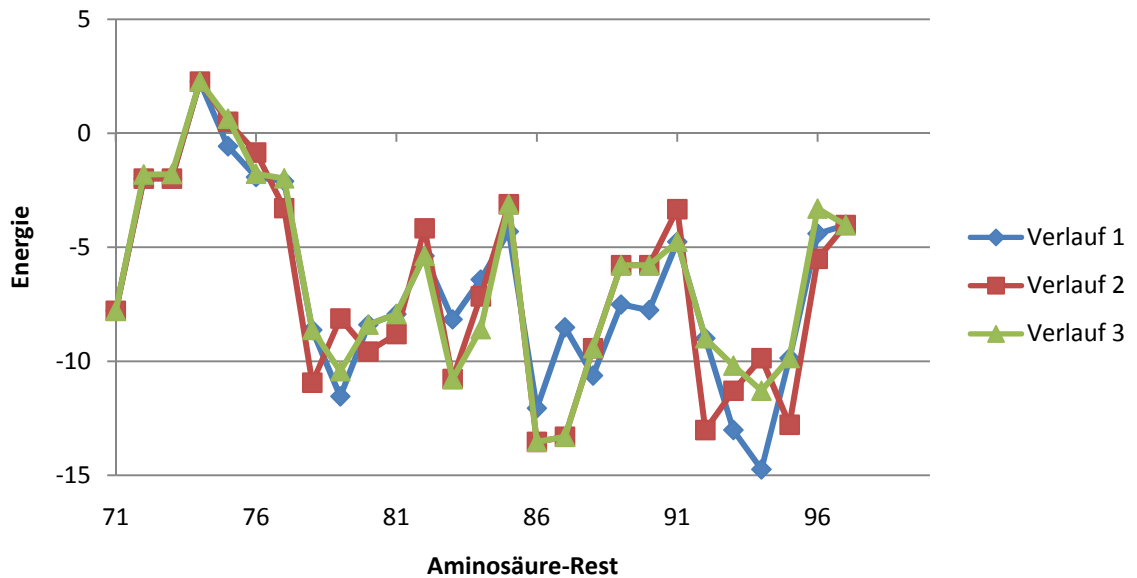


Abbildung 39: Energieprofil der intermediären Helix C Struktur aller 3 Verläufe. Dabei ist der statische Teil mit einbezogen.

Ein weiterer Ansatz zur Untersuchung des Widerspruchs zwischen den End-End Abständen und den Energien der intermediären Strukturen wird der Kraftverlauf über die Zeit mit den jeweiligen Strukturen in Abbildung 40 betrachtet. So zeigt sich der Kraftunterschied zwischen den intermediären Formen deutlich, sowie die Unterschiede in den Strukturen.

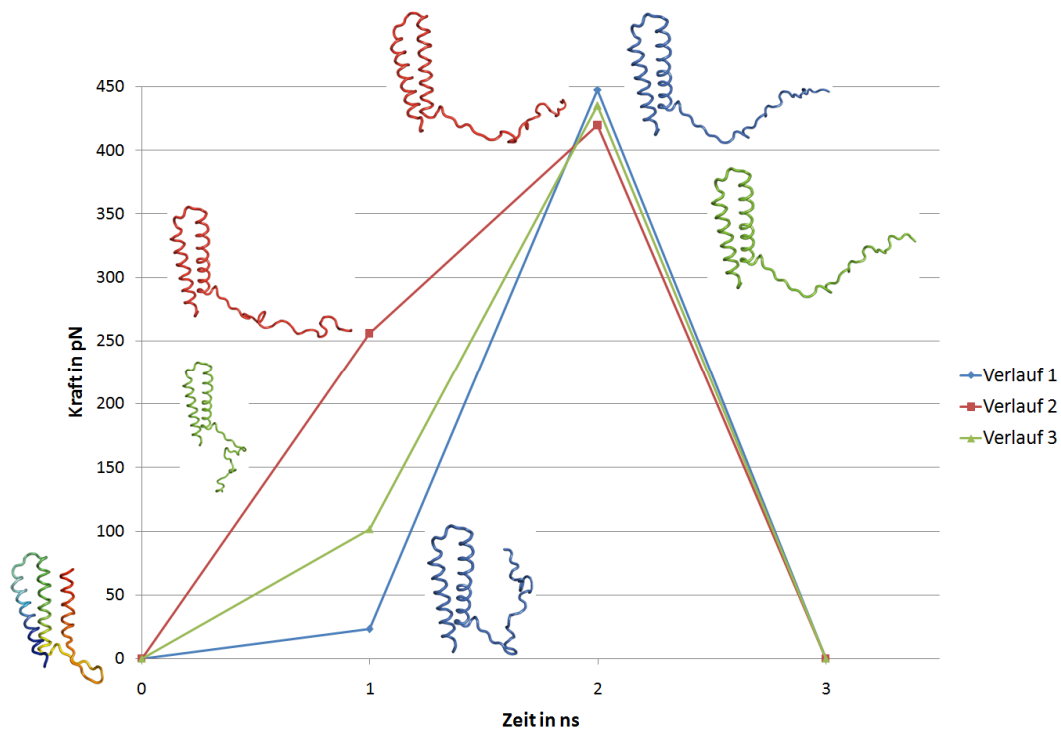


Abbildung 40: Kraftverlauf über die Entfaltungsdauer der Helix C mit statischem Teil

Die dargestellten Strukturen zeigen zwei verschiedene Verläufe. Zum einen in Verlauf 2 spreizt sich Helix C von dem angefügten statischen Teil ab. Zum anderen in Verlauf 1 und 3 beginnt sich die Helix zuerst zu entwinden. Dabei fällt jedoch auf, dass die erste Konformation in Verlauf 3 sich gleichzeitig in die entgegengesetzte Richtung dreht und entfaltet. Diese Drehung spiegelt sich in dem höheren Kraftausschlag wieder, als die gemessene Kraft in Verlauf 1. Jedoch spiegelt damit Verlauf 3 nicht die Kraftrichtung wieder, in welche die Helix gezogen wird.

In der Simulation wird der erste Rest des statischen Teils als der Punkt betrachtet, welcher an dem Objektisch befestigt wird und an den letzten Rest der Helix C wird die Cantilever angebracht. Dabei zeigt sich, dass der Verlauf 3 nicht der angelegten Kraftrichtung folgt bzw. nicht richtig auf die Richtung der

Kraft antwortet. Denn so zeigt sich in Abbildung 41, dass die intermediäre Form des Verlaufs 3 unterhalb des Objektisches liegt, was nicht möglich ist. Aus diesem Grund wird der Verlauf 3 für die weitere Betrachtung ausgeschlossen.

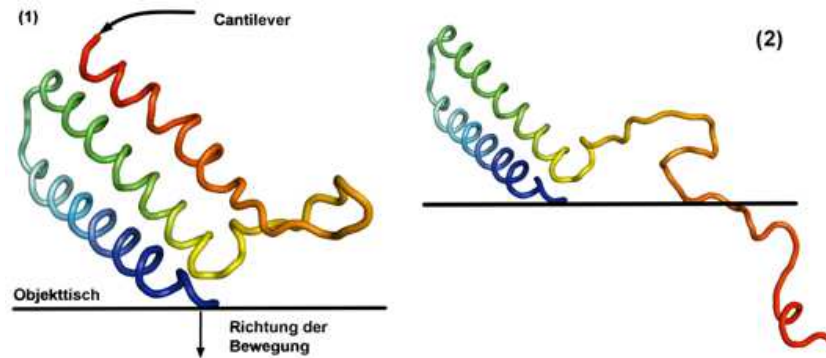


Abbildung 41: Darstellung unmöglicher Entfaltung der Helix C, (1) zeigt Helix C in der nativen Form mit Befestigung an Objektisch und Cantilever, (2) intermediäre Form von Verlauf 3. Es zeigt sich, dass die Struktur dabei mit dem Objektisch kollidiert und deshalb nicht möglich ist.

Mittels der Wahrscheinlichkeit der Konformationen pro Zeitschritt in Tabelle 15 zeigt sich, dass beide Verläufe ungefähr gleich auftreten. Die Wahrscheinlichkeiten der einzelnen Konformation sind jedoch wieder sehr gering, weshalb die Aussage, dass zwei Entfaltungswege existieren nicht eindeutig bestätigt werden kann. Diese Aussage ist aber durch Experimente nachgewiesen [Dressel, 2008].

Tabelle 15: Häufigkeiten der Konformation pro Verlauf für Helix C mit statischem Teil

Zeit in ns	Verlauf 1	Verlauf 2
1	0,4 %	0,37 %
2	1,87 %	1,63 %

6.2.2 Entfaltung von Helix C ohne statischen Teil

Für die Helix C ohne statischen Teil ergab sich die in Abbildung 42 dargestellte Kraft-Abstandskurve. Dabei weisen Verlauf 1 und 3 ein ähnliches Kraftmaximum bei ca. 150 pN auf und besitzen einen entfalteten End-End Abstand von ca. 5,4 nm. Verlauf 2 hingegen besitzt ein geringeres Kraftmaximum sowie einen geringeren End-End Abstand. Desweiteren besitzen Verlauf 1 und 3 nur eine intermediäre Form, Verlauf 2 hingegen besitzt zwei. Dabei liegt die zweite intermediäre Form in Bezug auf den End-End Abstand und die Kraft nahe am entfalteten Zustand. Nahe an der zweiten intermediären Form des Verlaufs 2, liegt die intermediäre Form des Verlaufs 1, weshalb diese in der Darstellung durch Verlauf 2 überdeckt wird. Die intermediäre Form des Verlaufs 3 befindet sich unterhalb der ersten intermediären Form des Verlaufs 2.

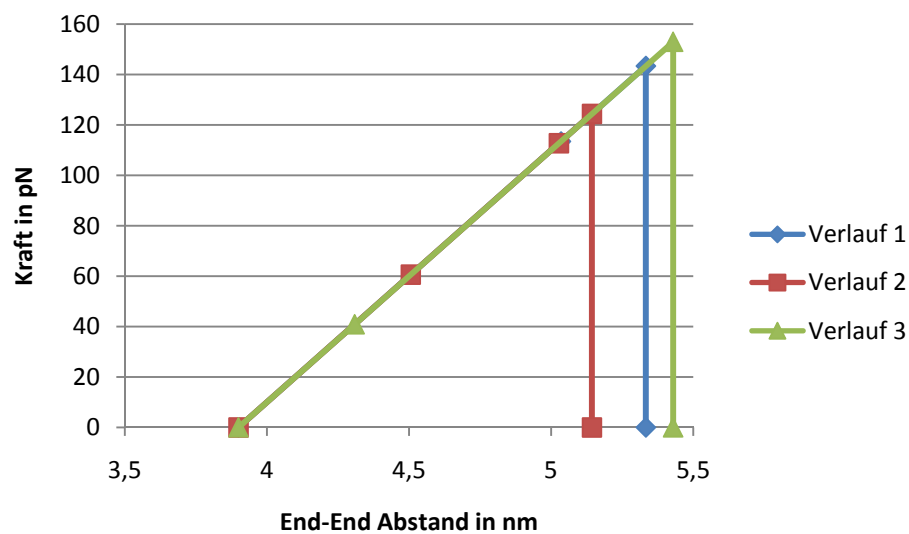


Abbildung 42: Kraft-Abstandskurve der Helix C ohne statischen Teil

Die zeitliche Änderung der Energie im Prozess der Entfaltung, dargestellt in Abbildung 43, zeigt nur geringe Abweichungen zwischen den einzelnen Verläufen. Bei Verlauf 2 ist jedoch von der zweiten zur dritten Nanosekunde ein geringer Abfall der Energie zu erkennen. Die geringere Energie würde bedeuten, dass sich Aminosäuren innerhalb der Struktur wieder näher kommen und damit auch der End-End Abstand sich verringern müsste. Dies ist allerdings, wie die Kraft-Abstandskurve zeigt, nicht der Fall. Das Gegenteil tritt ein, der End-End Abstand vergrößert sich.

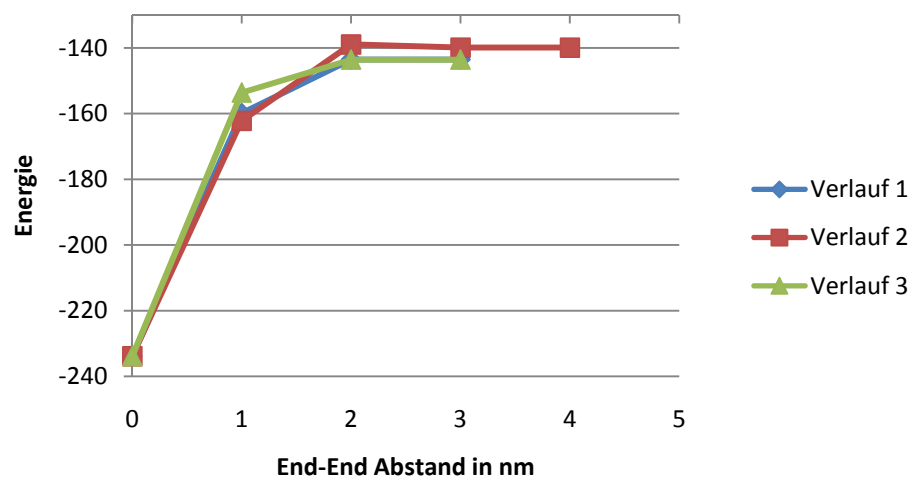


Abbildung 43: Energetischer Entwicklung der Entfaltung von Helix C ohne statischen Teil für alle 3 Verläufe

Um diesen Umstand zu verstehen bzw. den Energien deren jeweilige Struktur zuzuordnen, zeigt Abbildung 44 den Kraftverlauf über die Entfaltungsdauer hinweg mit den jeweiligen Strukturen. Dabei zeigt sich ähnlich wie schon bei der Helix B mit statischem Teil, dass jeweils ein intermediärer Zustand den intermediären Zuständen der anderen beiden Verläufe ähnelt. Hierbei gleichen sich Zwischenstruktur und die aufgewendete Kraft des Verlaufs 3 der ersten Zwischenstruktur des Verlaufs 2. Desweiteren weisen die intermediäre Form von Verlauf 1 mit der intermediären Form des Verlaufs 2 strukturelle Ähnlichkeiten, sowie ähnliche Kräfte auf. Jedoch scheinen sich die entfalteten Zustände von Verlauf 1 und 3 stark von dem des Verlauf 2 zu unterscheiden.

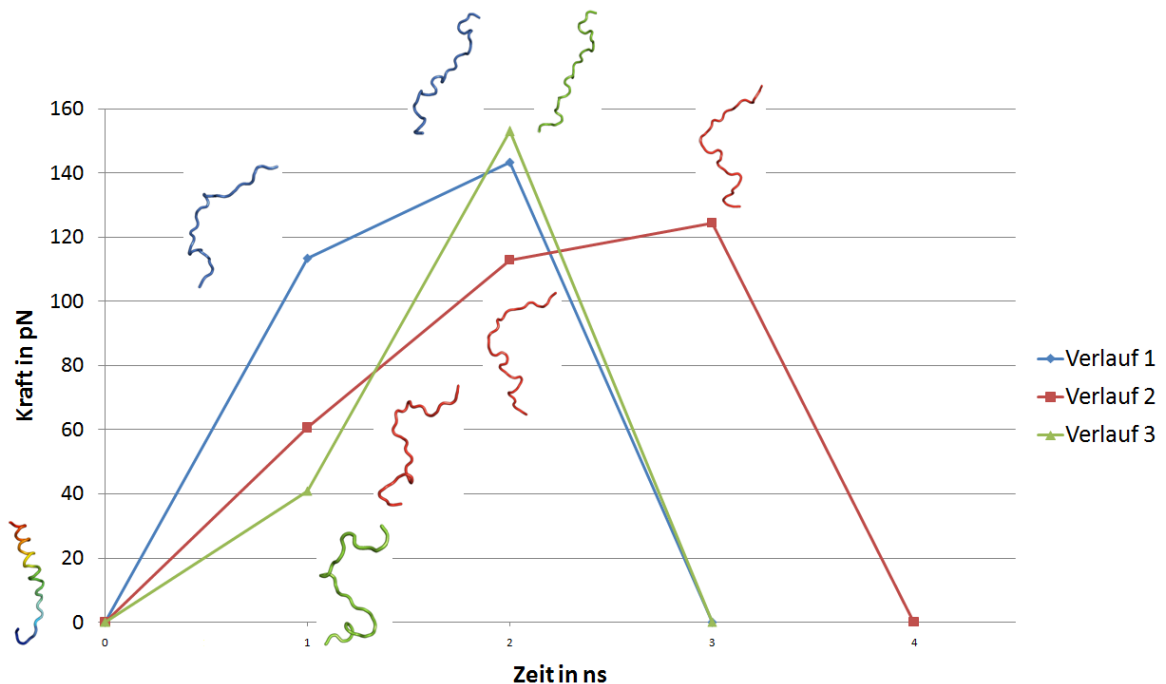


Abbildung 44: Kraftverlauf über die Entfaltungsdauer mit Strukturen der Helix C ohne statischen Teil

Für die zueinander passenden intermediären Formen, sowie die entfalteten Zustände wurden RMSD-Werte bestimmt. Die Ergebnisse sind in Tabelle 16 zu sehen. Dabei zeigt sich wie vermutet, dass sich die intermediäre Form des Verlaufs 3 mit der ersten intermediären Form des Verlaufs 2 ähnelt. Auch die Vermutung mit der Ähnlichkeit der zweiten intermediären Form bestätigt sich, jedoch ist dabei der Unterschied im Vergleich zur ersten Zwischenstruktur viel größer. Auch der entfaltete Zustand des Verlaufs 1 und 2 sind fast identisch. Der Unterschied zwischen den beiden Strukturen und der entfalteten Form von Verlauf 2 ist dabei größer und sollte nicht gleichgesetzt werden mit den anderen beiden. Somit scheint, dass sich Helix C ohne einen statischen Teil in zwei unterschiedliche Formen entfalten kann und dabei zwei Entfaltungswege genutzt werden.

Grundsätzlich zeigt sich wieder, dass die Kraft mit der Helix C alleinstehend entfaltet wird geringer ist, als mit dem statischen Teil. Die Kraft verhält sich dabei ähnlich wie bei der Entfaltung von Helix C.

Tabelle 16: RMSD-Werte der Strukturen für Helix C ohne statischen Teil

	Verlauf 1	Verlauf 2	Verlauf 3
Erste Intermediäre Form			
Verlauf 3	6,996 Å	3,843 Å	-
Zweite intermediäre Form			
Verlauf 1	-	6,003 Å	6,996 Å
Entfaltete Form			
Verlauf 1	-	1,932	0,001
Verlauf 2	1,932	-	4,222 Å
Verlauf 3	0,001 Å	4,222 Å	-

Betrachtet man noch die Häufigkeiten, in Tabelle 17, wie oft die Konformationen pro Verlauf vorkommen, so zeigt sich, dass dies eine sehr geringe Anzahl ist. Damit ist keine exakte Auskunft darüber möglich, wie das Entfaltungsverhalten der alleinstehenden Helix C ist.

Tabelle 17: Häufigkeit der Konformationen pro Zeitschritt in allen drei Verläufen. Dabei zeigen alle ein geringes, jedoch gleich häufiges Auftreten.

Zeit in ns	Verlauf 1	Verlauf 2	Verlauf 3
1	0.17 %	0.13 %	0.13 %
2	0.5 %	0.5 %	0.47 %
3	-	1.7 %	-

7 Auswertung

Betrachtet man die Kraftverläufe bei der Entfaltung der jeweiligen Helix, erkennt man die für diese SM-Methode typischen Anstiege der Kraft und einen abrupten Abfall, nachdem die jeweilige Helix entfaltet wurde.

Vergleicht man die drei Helices in Abbildung 45 miteinander, wobei Helix B und C einen statischen Teil besitzen, so weisen alle drei Helices unterschiedliche Kraftmaxima. Diese enormen Kraftunterschiede ergeben sich dabei durch die statischen Teile, sowie der Länge des entfalteten Teils und der Lage zum, an der simulierten Oberfläche befestigten, N-terminalen Ende.

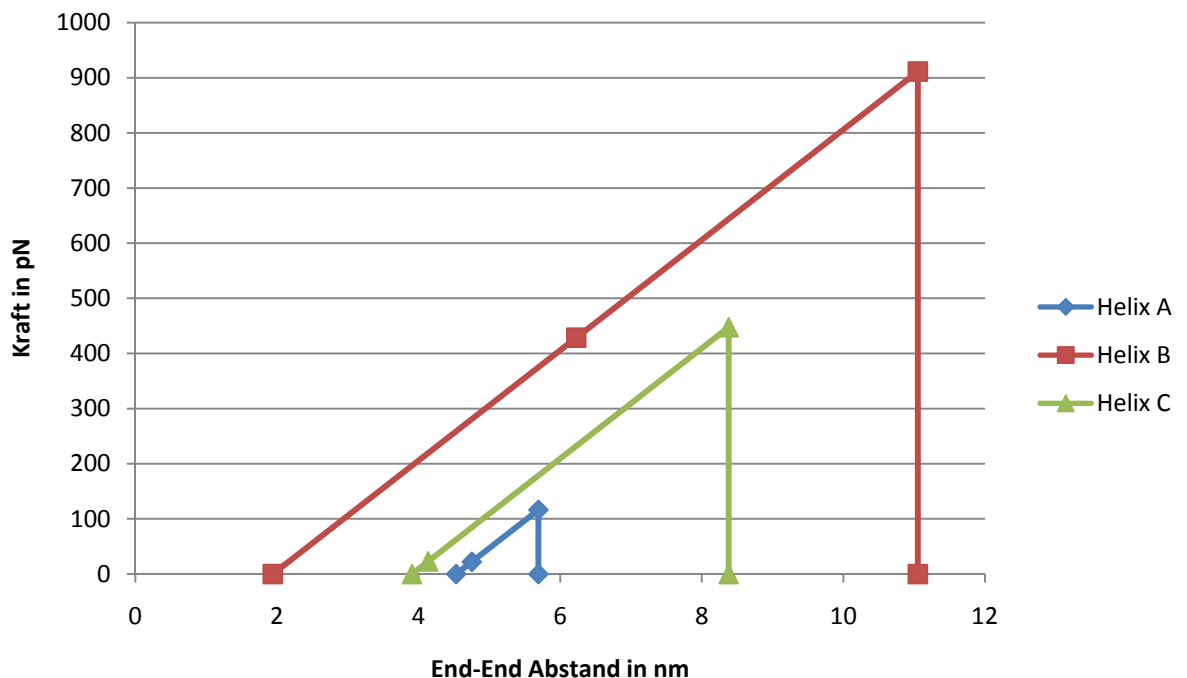


Abbildung 45: Kraft-Abstandskurven der Helix A, B und C. Dargestellt sind die häufigsten Verläufe pro Helix, dabei zeigt sich der Einfluss der statischen Teile sowie die Orientierung der entfaltbaren Teile.

Mittels der statischen Teile zeigt sich der Einfluss der Membran, welche die Helices durchziehen. Durch die Nähe der entfaltbaren Helices zu den anderen, treten diese ebenfalls in Wechselwirkung, wodurch die Stabilität erhöht wird und deshalb mehr Kraft erforderlich ist um Helix C und B zu entwinden.

Helix A lässt sich dabei mit weniger Kraft entfalten, da sie keinen weiteren statischen Teil besitzt. Der Grund für den sehr hohen Kraftausschlag von Helix B im Vergleich zu Helix A und C, ergibt sich dabei durch die Länge des Abschnittes, aus welchem der entfaltbare Teil besteht, sowie der Entfernung vom N-terminalen Ende, welcher als Bezugspunkt für den End-End Abstand diente. Der Abstand von Helix C ist geringer, da diese Helix kürzer ist als Helix B und der End-End Abstand nicht so groß ist, wie bei Helix B.

Die Häufigkeiten der jeweiligen Konformation pro Entfaltung ist in alle Simulationen sehr gering. Dies zeigt, dass es zu viele sehr ähnliche Möglichkeiten für jede der gewählten Helix gibt. Dadurch ist die Wahl der Fragmente-Namen, sowie die Anzahl der Fragmente-Schwerpunkte, welche für die Konstruktion einer Struktur genutzt wird, ausschlaggebend dafür ist, wie verwertbar das Ergebnis ist. Hierbei muss zwischen zu vielen Möglichkeiten und zu wenigen Möglichkeiten abgewogen werden. Zu viele, wie in dem Fall der 3 Helices ergeben sehr viele teils sehr unterschiedliche Entfaltungsverläufe. Jedoch würden zu wenige, d. h. durch eine Gruppierung weniger ähnlicher End-End Abstände, das Entfaltungsverhalten wiederum zu stark in eine Richtung gedrängt, was wiederum nicht den realen Bedingungen entspricht.

8 Zusammenfassung und Ausblick

Mittels des erstellten Programm-Paketes können mögliche Konformationen von Proteinabschnitten bzw. einem gesamten Protein modelliert werden und mittels dieser möglichen Konformation kann die mechanische Entfaltung einzelner Abschnitte simuliert werden. Dabei können die jeweiligen Entfaltungsschritte vollständig nachvollzogen werden. Jedoch kann das Programm verbessert werden. Wie die Entfaltung der Helix C zeigt, existieren Konformation, welche in eine falsche Richtung verlaufen, z.B. in den Objekttisch, was real nicht möglich ist. Durch diese Einschränkung des möglichen Raumes lassen sich die möglichen Konformationen einschränken und das Ergebnis verbessern. Desweiteren könnten die Ergebnisse durch eine Verringerung der Fragment-Schwerpunkte pro Fragment-Namen verbessert werden, da dies zu einem größeren Unterschied zwischen den Fragmenten führen würde. Dies würde die Häufigkeiten der genutzten Konformationen erhöhen.

Die Ergebnisse der drei Helices zeigen, dass die gewählte Methode zur Auswertung der erzeugten Mittelwerte noch nicht optimal ist und deshalb überarbeitet werden sollte bzw. eine neue Methode ausgewählt werden sollte, um eindeutiger Ergebnisse zu liefern. Die hohe Varianz der aufgetretenen Strukturen gibt keinen sicheren Aufschluss über das Entfaltungsverhalten. Diese hohe Varianz könnte durch die schon angesprochenen Methoden verringert werden bzw. durch höhere Durchläufe oder eine andere Auswertung.

Grundsätzlich zeigen die Ergebnisse, dass der angefügte statische Teil den Einfluss der Membran, durch höhere Kraftausschläge als bei den Helices ohne statischen Teil, gut simuliert.

Literaturverzeichnis

[Andreopoulos & Labudde, 2011]

Andreopoulos, Bill; Labudde, Dirk (2011): *Efficient unfolding pattern recognition in single molecule force spectroscopy data*. Algorithms for Molecular Biology. 6: Seite 1-6

[Bruno et al, 2005]

Bruno, S.; Zuccheri, G.; Baschieri, P. (2005): *Protein Unfolding and Refolding Under Force: Methodologies for Nanomechanics*. ChemPhysChem. 6: Seite 29-34

[Bryngelson et al, 1995]

Bryngelson, Joseph D.; Onuchic, José Nelson; Socci, Nicholas D.; Wolynes, Peter G. (1995): *Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis*. PROTEINS: Structure, Function, and Genetics. 21: Seite 167-195

[Cieplak et al, 2006]

Cieplak, M; Filipek, Slawomir; Janovjak, Harald; Krzyśko, Krystiana A. (2006): *Pulling single bacteriorhodopsin out of a membrane: Comparison of simulation and experiment*. Biochimica Et Biophysica Acta, 1758: Seite 537-544

[Deniz et al, 2008]

Deniz, Ashok. A.; Samrat, Mukhopadhyay; Lemke, Edward. A. (2008): *Single-molecule biophysics: at the interface of biology, physics and chemistry*. Journal of The Royal Society Interface. 5: Seite 15-45

[Dressel, 2008]

Dressel, Frank (2008): *Sequenz, Energie , Struktur – Untersuchungen zur Beziehung zwischen Primär- und Tertiärstruktur in globulären und Membran-Proteinen*, 87 Seiten Dresden, Technische Universität Dresden, Mathematische-Naturwissenschaftliche Fakultät, Dissertation

[Haynie & Freire, 1993]

Haynie, Donald T.; Freire, Ernesto (1993): *Structural Energetics of the Molten Globule State*. PROTEINS: Structure, Function, and Genetics. 16: Seite 115-140

[Heinke, 2010]

Heinke, Florian (2010): *Energieprofilbasierende Analysemethoden von Proteinfamilien*, 83 Seiten Mittweida, Hochschule Mittweida, Mathematik/Naturwissenschaften/Informatik, Bachelorarbeit

[Ivarsson et al, 2008]

Ivarsson, Ylva; Travaglini-Allocatelli, Carlo; Brunori, Maurizio; Gianni, Stefano (2008): *Mechanisms of protein folding*. European Biophysics Journal. 37: Seite 721-728

[Jaenicke, 1984]

Jaenicke, Rainer (1984); *Proteinfaltung und Proteinassoziation*. Angewandte Chemie. 6: Seite 385-454

[Janshoff et al, 2000]

Janshoff, Andreas; Neitzert, Marcus; Oberdörfer, York; Fuchs, Harald (2000): *Force Spectroscopy of Molecular Systems – Single Molecule Spectroscopy of Polymers and Biomolecules*. Angewandte Chemie. 39: Seite 3212-3237

[King et al, 2010]

King, William T.; Su, Meihong; Yang, Guoliang. (2010): *Monte Carlo simulation of mechanical unfolding of proteins based on a simple two-state model*. International Journal of Biological Macromolecules. 46: Seite 159-166

[Löffler et al, 2007]

Löffler, Georg; Petrides Petro E.; Heinrich, Peter C. (2007): *Biochemie und Pathobiochemie*. 8. Auflage: Springer

[Makarov et al, 2001]

Makarov, Dmitrii E.; Hansma, Paul K.; Metiu, Horia (2001): *Kinetic Monte Carlo simulation of titin unfolding*. Journal of Chemical Physics. 114, 21: Seite 9663-9673

[Onuchic et al, 1997]

Onuchic, José Nelson; Luthey-Schulten, Zaida; Wolynes, Peter G. (1997): *Theory of Protein Folding: The Energy Landscape Perspective*. Annual Reviews A nonprofit scientific publisher. 48: Seite 545-600

[Schildbach, 2010]

Schildbach, Stefan (2010): *Aufbau eines Servers für die Energieprofilbasierte Suche*, 53 Seiten Mittweida, Hochschule Mittweida, Mathematik/Naturwissenschaften/Informatik, Bachelorarbeit

[Schlierf & Rief, 2009]

Schlierf, Michael; Rief, Matthias (2009): *Überraschend einfache Einzelmolekülmechanik der Faltung von Proteinen*. Angewandte Chemie. 121: Seite 835-837

[Voet et al, 2002]

Voet, Donald; Voet, Judith G.; Pratt, Charlotte W. (2002): *Lehrbuch der Biochemie*. 1. Auflage: Willey-VCH.

[Ye & Godzik, 2003]

Ye, Yuzhen; Godzik, Adam (2003): *Flexible structure alignment by chaining aligned fragment pairs allowing twists*. Bioinformatics. 19: 246-255

[Zhang et al, 2009]

Zhang, Jian; Li, Wenfei; Wang, Jun. (2009): *Protein folding simulations: From coarse-grained model to all-atom model*. IUBMB Life. 61: Seite 627-643

[Zwanzig, 1997]

Zwanzig, Robert (1997): *Two-state models of protein folding kinetics*.
Proceedings of the National Academy of Sciences of the United States of
America. 94: Seite 148-150

[URL 1]

(30.06.2012) Christine Weinl: Vorlesung „Zellbiologie für Bioinformatiker“

URL: <http://www.uni-tuebingen.de/uni/kxm/Courses/documents/membranen.pdf>

[URL 2]

(06.07.2012) Unbekannter Autor: *Aminoacid Venn Diagram* - Dynamique des
Systèmes et Interactions des Macromolécules Biologiques

URL: http://www.dsimb.inserm.fr/~debrevern/VENN_DIAGRAM/aa_venn_diagram.png

[URL 3]

(10.07.2012) Unbekannter Autor: Ramachandran Plot

URL: http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html

[URL 4]

(05.06.2012) Unbekannter Autor: TMDET: Detection of transmembrane regions

URL: <http://tmdet.enzim.hu/?go=form>

[URL 5]

(22.06.2012) Max Planck Institute of Biochemistry: Halobacterium signal
transduction network URL:

http://www.biochem.mpg.de/en/eg/oesterhelt/web_page_list/Org_Hasal/absatz_6_bild.gif

Anhang

Sämtliche Ergebnisse über die Entfaltungssimulation der drei Helices, sowie die erstellten Dateien und Quellcodes der erstellten Programme befinden sich auf der CD „Anhang“.

Desweiteren befindet sich auf der CD eine digitale Kopie dieser Arbeit.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, den 13.08.2012

Eric Zuchantke